Real Time Sequential Detection for Diffusion Signals

by

J.S. Baras and A. LaVigna

Report Documentation Page

1. REPORT DATE **1986**	2. REPORT TYPE	3. DATES COVERED **00-00-1986 to 00-00-1986**

4. TITLE AND SUBTITLE **Real Time Sequential Detection for Diffusion Signals**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,Electrical Engineering Department,College Park,MD,20742**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **68**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

REAL TIME

SEQUENTIAL DETECTION FOR DIFFUSION SIGNALS

by

John S. Baras*   and   Anthony LaVigna**

Electrical Engineering Department
and
Systems Research Center

University of Maryland
College Park, Maryland 20742

Abstract

We analyze sequential detection for diffusion type signals both in the fixed probability of error formulation and in the Bayes Formulation. We show that the optimal strategy in both cases is of the threshold type with explicitly computable thresholds. We provide efficient numerical schemes for computing approximations to the likelihood ratio and provide an implementation via a special purpose VLSI processor for real time processing in the scalar diffusion case. Finally, we describe the DELPHI expert system which is under development. Its purpose is to provide an integrated system level design tool for sequential real time detection and estimation.

Introduction

In the present paper we analyze in detail the sequential detection problem for diffusion process signals. In particular we address questions related to optimal strategies, numerical computation of the sufficient statistics, real-time architectures for implementing the resulting algorithms, and describes a sophisticated system level design tool: the DELPHI system. The techniques presented here can be extended to apply to more general signal, observation models such as multi-dimensional diffusions, jump point process models and mixed models. We shall not address them here.

While Gaussian signal processing theory and design have reached a certain degree of completeness, the corresponding developments for non-Gaussian signal processing are unsatisfactory from the designer's point of view. The main reason is that no systematic effort has been undertaken to transform the theoretical advances in non-Gaussian detection and estimation theory into practical design methods. In the meantime the engineering specifications on signal processors are becoming increasingly more demanding, resulting in an unavoidable ascending degree of complexity.

Recent progress in stochastic processes, and in particular in *non-Gaussian detection* and *nonlinear filtering theory*, hold great promise. It is well known that every signal detection and classification problem requires the solution of some underlying estimation problem (Kailath [1969]). Despite the fact that this has been known for many years, its impact on the resolution of complex signal processing problems in non-Gaussian environments has been very limited (see the references of (Kailath

1

[1969]) for some ingenious applications). The primary reason for the limited applicability of the powerful machinery of nonlinear estimation in signal detection and classification problems has been the formidable computational complexity called for by these theories: at least a *stochastic nonlinear* partial differential equation has to be solved on-line. However, recently three events have caused a serious reexamination of these issues. First, recent progress in nonlinear estimation succeeded in replacing the fundamental stochastic nonlinear partial differential equation with a *linear non-stochastic* partial differential equation, (Davis [1981]) the so called robust version of the Zakai equation. Second, the advent of *systolic VLSI arrays* (as well as other special purpose VLSI array processors) led to the proposal of a design of a VLSI processor that could implement the Zakai equation in real time (Baras [1981]). Third, the maturation of *artificial intelligence* has provided a plethora of design tools whose impact on other disciplines has not even been tested yet.

The nonlinear filtering problem, in its various manifestations (depending on signal and observation models), is at the heart of many interesting and significant signal processing problems. We have identified, among others, digital signal processing (such as pulse amplitude modulation, delta modulation, adaptive delta modulation, and speech processing), direction finding receivers, digital phase lock loops, adaptive sonar and radar arrays, sequential detection, simultaneous detection and estimation, sensor scheduling in data fusion problems, adaptive stochastic control, stochastic control with partial observations, and nonlinear observers.

We next offer some justification for the selection of the research problem and technical approach. First regarding VLSI architectures, we note that recently a lot of attention and research effort has been devoted to VLSI architectures for linear signal processing schemes (Kailath Whitehead]). However, the need for special architectures is more dramatic in the case of nonlinear signal processors, such as needed in non-Gaussian problems. To focus, we recall that the sequential detection problem or the

nonlinear filtering problem, calls for a processor that operates on data in "real time". What is meant by "real time"? Since it is clear that we are led to some sort of digital circuit implementation, real time means that the allowable sampling period for the observed data must be larger that the computational cycle for the digital device implementing the algorithm. In the case of the nonlinear filtering problem, the theoretical analysis produces an algorithm that calls for the solution of a stochastic partial differential equation. Hence a real time processor based on this algorithm must provide fast approximations to the solution of this equation. To a great extent our work has been initiated by the desire for developing a theory and design to account for real time implementation constraints for non-Gaussian detection and non-linear filtering.

The structure of this paper is as follows.

In Section 1, we provide a detailed analysis of the sequential detection problem for diffusion signal and observation models. Both the Bayesian and fixed probability of error formulations are examined. We show that in both cases the optimal strategy is of *threshold type* with thresholds that can be computed explicitly.

In Section 2, we provide a detailed treatment of the numerical analysis of the *Zakai equation* using semigroup methods. A general approximation theorem is presented. This theorem is intended to be used to check convergence of individual approximation schemes.

In Section 3, we describe a VLSI architecture that can implement the algorithms presented in Section 2 in *real time*. Initial estimates indicate that for 100 mesh points, we can perform 5000 calculations per second.

In Section 4, we describe the DELPHI expert system, which is under development at Maryland. It affords an engineer a sophisticated and " user friendly" design tool for real-time circuit design iin non-Gaussian detection and estimation problems.

Finally, we discuss briefly some future direction of the work described here.

1. Sequential Detection of Diffusion Signals

1.1 Introduction

In this chapter, we consider the binary sequential hypothesis testing problem. Here, we are given an $\mathbb{R}^n$-valued signal process $\{x_t, \ t \geq 0\}$ which satisfies the stochastic differential equation

$$dx_t = f(x_t) \, dt + g(x_t) \, dw_t$$

$$x_0 = \xi$$

(1.1)

where $\{w_t, \ t \geq 0\}$ is an $\mathbb{R}^m$-valued standard Brownian motion. However, we cannot observe $\{x_t, \ t \geq 0\}$ directly, instead we only observe the increments dy_t of an $\mathbb{R}^p$-valued stochastic process $\{y_t, \ t \geq 0\}$. Under each hypothesis the observed data is the output of a stochastic differential equation, i.e.,

$$\text{Under } H_1: \qquad dy_t = h(x_t) \, dt + dv_t$$

$$\text{Under } H_0: \qquad dy_t = dv_t$$

(1.2)

where $\{v_t, \ t \geq 0\}$ is an $\mathbb{R}^p$-valued standard Brownian motion which is independent of $\{w_t, \ t \geq 0\}$.

We assume that for all x, y in $\mathbb{R}^n$, the functions f, g, and h satisfy the *Lipschitz* condition,

$$\|f(x) - f(y)\| + \|g(x) - g(y)\| + \|h(x) - h(y)\| \leq K\|x - y\| \qquad (1.3)$$

and the *growth* condition

$$\|f(x)\|^2 + \|g(x)\|^2 + \|h(x)\|^2 \leq K^2(1 + \|x\|^2). \qquad (1.4)$$

These conditions guarantee that the stochastic differential equations in (1.1) *have unique continuous strong solutions* (Arnold [1974]).

Data is observed continuously starting at an initial time which is taken for convenience to be zero. At each time $t > 0$, the decision-maker can either declare one of the hypotheses to be true or continue collecting data. The decision-maker selects his decision based on the data collected up to time t, so as to minimize an appropriate cost function.

We will present both the fixed probability of error and the Bayesian formulations for this problem. In both formulations, we have a measurable space $(\Omega, \mathcal{F})$, on which we are given two probability measures P_0, P_1, and the random process $\{y_t, \ t \geq 0\}$. When hypothesis H_0 (respectively H_1) is valid the statistics of the observed process $\{y_t, \ t \geq 0\}$ are governed by measure P_0 (respectively P_1).

The difference between these formulations is how they prescribe the cost function to be minimized. For each formulation, we show that the appropriate cost function is minimized by a threshold policy. The essential difference between these formulations is how they prescribe the thresholds.

More precisely, a decision policy involves the selection of a termination time τ, and of a binary valued decision δ. If $\delta = 1$, we shall accept hypothesis H_1; if $\delta = 0$ we shall accept hypothesis H_0. Let $\mathcal{F}_t^y$ denote the σ-algebra generated by $\{y_s, \ s \leq t\}$.

Definition 1.1.1. An *admissible decision policy* is any pair $u = (\tau, \delta)$ of RV's where τ is an $\mathcal{F}_t^y$-stopping time, and δ is an $\mathcal{F}_\tau^y$-measurable $\{0, 1\}$-valued RV. The collection of all admissible decision policies will be denoted by $\mathcal{U}$.

Definition 1.1.2. A policy u in $\mathcal{U}$ is a *threshold policy* or *of threshold type* if there exists constants A and B, with $0 < A \leq 1 \leq B < \infty$ and $A \neq B$, such that

$$\tau = \inf(t \geq 0 \mid \Lambda_t \notin (A, B)) \tag{1.5}$$

$$\delta = \begin{cases} 1, & \Lambda_\tau \geq B \\ 0, & \Lambda_\tau \leq A \end{cases} \tag{1.6}$$

Here Λ_t is the likelihood ratio associated with this problem, namely

$$\Lambda_t = \exp\left(\int_0^t \hat{h}_s^T \, dy_s - \frac{1}{2} \int_0^t \|\hat{h}_s\|^2 \, ds\right) \tag{1.7}$$

where T denoted transpose, and

$$\hat{h}_t = E_1\left(h(x) \mid \mathcal{F}_t^y\right). \tag{1.8}$$

This chapter is organized as follows. First, we present general results on threshold policies. Then, we show how threshold policies solve the fixed probability of error problem. Then, we show that, with the cost function given in (1.42), threshold policies solve the Bayesian problem. Finally, we state the generalization to the case where H_0, like H_1, includes a function of a signal process $\{x_t^0, \; t \geq 0\}$. In both of these problems, we show how to find the optimal threshold policy.

Throughout this chapter, we make the following technical assumptions.

(T1) $E_i\left(|h(x_t)|\right) < \infty, \; t \geq 0.$

(T2) $P_i\left(\int_0^\infty \|\hat{h}_s\|^2 \, ds = \infty\right) = 1.$

(T3) $E_i\left(\int_0^t \|\hat{h}_s\|^2 \, ds\right) < \infty, \; 0 \leq t < \infty.$

where $\hat{h}_t$ is defined in (1.8).

1.2 Threshold Policies

For each u in $\mathcal{U}$, let $\alpha(u)$ and $\beta(u)$ be the false alarm and miss probabilities associated with u, respectively, i.e.

$$\alpha(u) = P_0(\delta = 1) \qquad \beta(u) = P_1(\delta = 0). \tag{1.9}$$

Throughout this chapter α and β will be positive *constants* that satisfy the inequality $\alpha + \beta < 1$.

A threshold policy u in $\mathcal{U}$ will be described in the form (1.5)–(1.8) and will be identified with the threshold constants (A, B). Let Γ be the collection of all threshold policies in $\mathcal{U}$.

We will now prove some results about threshold policies. These proofs are taken from (Liptser & Shiryayev [1978, Chapter 17.6]).

Lemma 1.2.1. *For a threshold policy u in Γ,*

$$P_i(\tau < \infty) = 1, \qquad i = 0, 1. \tag{1.10}$$

Proof: We will show the result for $i = 1$ as a similar argument works for the case $i = 0$. Let σ_n be the sequence of $\mathcal{F}_t^y$-stopping times defined by

$$\sigma_n = \inf(t \geq 0 \mid \int_0^t \|\hat{h}_s\|^2 \, ds \geq n)$$

for each $n = 0, 1, \ldots$. The very definition of τ yields the inequalities

$$A \leq \exp(\int_0^{\tau \wedge \sigma_n} \hat{h}_s^T \, dy_s - \frac{1}{2} \int_0^{\tau \wedge \sigma_n} \|\hat{h}_s\|^2 \, ds) \leq B.$$

or equivalently

$$\log A \leq \int_0^{\tau \wedge \sigma_n} \hat{h}_s^T \, dy_s - \frac{1}{2} \int_0^{\tau \wedge \sigma_n} \|\hat{h}_s\|^2 \, ds \leq \log B. \tag{1.11}$$

Under P_1, it can be shown (Liptser & Shiryayev [1977]) that $\{y_t,\ t \geq 0\}$ satisfies the stochastic differential equation

$$dy_t = \hat{h}_t dt + d\overline{W}_t \tag{1.12}$$

with $\{\overline{W}_t,\ t \geq 0\}$ a standard Brownian motion, and upon substituting (1.12) into (1.11) we see that

$$\log A \leq \int_0^{\tau \wedge \sigma_n} \hat{h}_s^T\, d\overline{W}_s + \frac{1}{2} \int_0^{\tau \wedge \sigma_n} \|\hat{h}_s\|^2\, ds \leq \log B \tag{1.13}$$

From the definition of σ_n, $\int_0^{\sigma_n} \|\hat{h}_s\|^2\, ds = n$, from which it follows that

$$E_i \left(\int_0^{\tau \wedge \sigma_n} \hat{h}_s^T\, d\overline{W}_s \right) = 0$$

Now, by taking the expectation of (1.13) under P_1, we get from (1.14) that

$$
\begin{aligned}
E_1(\log \Lambda_{\tau \wedge \sigma_n}) &= E_1\left(\frac{1}{2} \int_0^{\tau \wedge \sigma_n} \|\hat{h}_s\|^2\, ds + \int_0^{\tau \wedge \sigma_n} \hat{h}_s^T\, d\overline{W}_s\right) \\
&= E_1\left(\frac{1}{2} \int_0^{\tau \wedge \sigma_n} \|\hat{h}_s\|^2\, ds\right) \leq \log B.
\end{aligned} \tag{1.14}
$$

Since (1.14) is true for all n and $\sigma_n \uparrow \infty$ it follows that

$$E_1\left(\frac{1}{2} \int_0^{\tau} \|\hat{h}_s\|^2\, ds\right) < \infty \tag{1.15}$$

whence

$$\infty > E_1\left(\frac{1}{2} \int_0^{\tau} \|\hat{h}_s\|^2\, ds\right) \geq E_1\left\{ 1_{(\tau=\infty)} \frac{1}{2} \int_0^{\infty} \|\hat{h}_s\|^2\, ds \right\}.$$

The conclusion

$$P_1(\tau = \infty) = 0$$

is now readily obtained via (T2). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Since Λ_t has a.s. continuous sample paths and since τ is a.s. finite, we can conclude that for threshold policies, Λ_τ takes on the values A or B (P_0- and P_1-a.s.).

Lemma 1.2.2. *For a threshold policy u in Γ with thresholds A and B,*

$$\alpha(u) = \frac{1 - A}{B - A}, \qquad \beta(u) = \frac{A(B - 1)}{B - A}. \tag{1.16}$$

Proof: From the comment following Lemma 1.2.1, we know that

$$\alpha(u) = P_0(\delta = 1) = P_0(\Lambda_\tau = B) \tag{1.17}$$

and

$$\beta(u) = P_1(\delta = 0) = P_1(\Lambda_\tau = A) \tag{1.18}$$

From Girsanov's Theorem (Liptser & Shiryayev [1977]) we see that

$$\frac{dP_1}{dP_0}(\mathcal{F}_\tau^y) = \Lambda_\tau,$$

hence

$$P_1(\Lambda_\tau = A) = E_0\big(1_{(\Lambda_\tau = A)}\Lambda_\tau\big) = A\, P_0(\Lambda_\tau = A) \tag{1.19}$$

similarly,

$$P_0(\Lambda_\tau = B) = E_1\big(1_{(\Lambda_\tau = B)}\Lambda_\tau^{-1}\big) = \frac{1}{B} P_1(\Lambda_\tau = B). \tag{1.20}$$

It also follows from the comment following Lemma 1.2.1 that

$$P_0(\Lambda_\tau = A) = 1 - P_0(\Lambda_\tau = B) \tag{1.21}$$

and

$$P_1(\Lambda_\tau = B) = 1 - P_1(\Lambda_\tau = A). \tag{1.22}$$

From (1.19)–(1.22) and some simple algebra, we get (1.16) $\qquad\square$

This result has several immediate consequences.

Corollary 1.2.3. *For any threshold policy u in Γ,*

$$\alpha(u) + \beta(u) < 1.$$

Proof: This follows from (1.16) above and from the fact that $0 < A \leq 1 \leq B < \infty$ with $A \neq B$.

Corollary 1.2.4. *Let u be a threshold policy in Γ with A and B defined by*

$$A = \frac{\beta}{1 - \alpha} \qquad B = \frac{1 - \beta}{\alpha} \tag{1.23}$$

where $\alpha + \beta < 1$, then

$$\alpha(u) = \alpha \qquad \beta(u) = \beta.$$

Proof: This is obtained by direct substitution of (1.23) into (1.16).

Lemma 1.2.5. *Let $u = (\tau, \delta)$ be any policy in $\mathcal{U}$ with $\alpha(u) + \beta(u) < 1$ and let*

$$\alpha := \alpha(u) \qquad \beta := \beta(u).$$

Define $u^* = (\tau^*, \delta^*)$ to be the threshold policy in Γ with parameters (A^*, B^*) that correspond to the pair (α, β) as defined in (1.23).

Then

$$E_0 \left(\int_0^{\tau^*} \|\hat{h}_s\|^2 \, ds \right) = 2w(\alpha, \beta) \tag{1.24}$$

$$E_1 \left(\int_0^{\tau^*} \|\hat{h}_s\|^2 \, ds \right) = 2w(\beta, \alpha) \tag{1.25}$$

where

$$w(x, y) := (1 - x) \log \frac{1 - x}{y} + x \log \frac{x}{1 - y}, \quad 0 < x, y < 1. \tag{1.26}$$

Proof: We will only show (1.24) since a similar argument works for (1.25). Let

$$L_t = \log \Lambda_t = \int_0^t \hat{h}_s^T \, dy_s - \frac{1}{2} \int_0^t \|\hat{h}_s\|^2 \, ds, \quad t \geq 0.$$

Consider the solution $g_i(x)$ of the boundary valued problem

$$\begin{cases} g_i''(x) + (-1)^{1+i} g_i'(x) = -2 \\ g_i(\log A) = g_i(\log B) = 0. \end{cases} \quad i = 0, 1 \tag{1.27}$$

on the interval $[\log A, \log B]$. Elementary calculations show that

$$g_1(x) = 2\left(\frac{B - ABe^{-x}}{B - A}\log\frac{B}{A} + \log A - x\right) \qquad (1.28)$$

$$g_0(x) = 2\left(\frac{B - e^x}{B - A}\log\frac{B}{A} - \log B + x\right), \qquad (1.29)$$

for $\log A \leq x \leq \log B$. From (1.15), (1.26), (1.28), and (1.29), we now see that

$$g_0(0) = 2w(\alpha, \beta) \qquad (1.30)$$

$$g_1(0) = 2w(\beta, \alpha). \qquad (1.31)$$

By applying Itô's rule to $g_0(L_{\tau \wedge \sigma_n})$, we get

$$g_0(L_{\tau \wedge \sigma_n}) = g_0(0) + \int_0^{\tau \wedge \sigma_n} g_0'(L_s)\hat{h}_s^T \, dy_s + \frac{1}{2}\int_0^{\tau \wedge \sigma_n} \left(g_0''(L_s) - g_0'(L_s)\right)\|\hat{h}_s\|^2 \, ds$$

$$= g_0(0) + \int_0^{\tau \wedge \sigma_n} g_0'(L_s)\hat{h}_s^T \, dy_s - \int_0^{\tau \wedge \sigma_n} \|\hat{h}_s\|^2 \, ds. \qquad (1.32)$$

Again from the definition of σ_n and the fact that $g_0'(x)$ is bounded for $x \in [\log A, \log B]$, we see that

$$E_0\left(\int_0^{\tau \wedge \sigma_n} g_0'(L_s)\hat{h}_s^T \, dy_s\right) = E_0\left(\int_0^{\tau \wedge \sigma_n} g_0'(L_s)\hat{h}_s^T \, dv_s\right) = 0$$

and consequently upon taking the expectation of (1.32) with respect to P_0 we get

$$E_0(g_0(L_{\tau \wedge \sigma_n})) = g_0(0) - E_0\left(\int_0^{\tau \wedge \sigma_n} \|\hat{h}_s\|^2 \, ds\right). \qquad (1.33)$$

Letting n go to infinity in (1.33), we finally obtain

$$0 = g_0(0) - E_0\left(\int_0^{\tau} \|\hat{h}_s\|^2 \, ds\right),$$

whence (1.24). The equality (1.25) can be established using similar methods. □

1.3 Fixed Probability of Error Formulation

Given $0 < \alpha, \beta < 1$ with $\alpha + \beta < 1$, let $\mathcal{U}(\alpha, \beta)$ be the set of all admissible policies u in $\mathcal{U}$ such that

$$\alpha(u) \leq \alpha \qquad \beta(u) \leq \beta$$

The fixed probability of error formulation to the sequential hypothesis testing problem requires the solution of the following.

Problem $(\mathcal{P}_F)$: *Find u^* in $\mathcal{U}(\alpha, \beta)$ such that for all u in $\mathcal{U}(\alpha, \beta)$,*

$$E_i\left(\int_0^\tau \|\hat{h}_s\|^2 \, ds\right) \geq E_i\left(\int_0^{\tau^*} \|\hat{h}_s\|^2 \, ds\right), \quad i = 0, 1. \tag{1.34}$$

Theorem 1.3.1. *If u^* is the threshold policy with constants (A^*, B^*) defined by*

$$A^* = \frac{\beta}{1 - \alpha}, \qquad B^* = \frac{1 - \beta}{\alpha},$$

then u^* solves problem $(\mathcal{P}_F)$.

Proof: From Lemma 1.2.2, it follows that $u \in \mathcal{U}(\alpha, \beta)$. Hence, we only need to show u^* is optimal is the sense of (1.33). To this end, let $u = (\tau, \delta)$ be any policy in $\mathcal{U}(\alpha, \beta)$. From Lemma 1.2.1 we see that

$$E_1\left(\frac{1}{2}\int_0^\tau \|\hat{h}_s\|^2 \, ds\right) = E_1\left(\int_0^\tau \hat{h}_s^T \, dy_s - \frac{1}{2}\int_0^\tau \|\hat{h}_s\|^2 \, ds\right)$$

$$= E_1(\log \Lambda_t)$$

$$= -E_1(\log \Lambda_\tau^{-1})$$

and from Jensen's inequality

$$E_1\left(\frac{1}{2}\int_0^\tau \|\hat{h}_s\|^2 \, ds\right) = -E_1(\log \Lambda_\tau^{-1}) \tag{1.35}$$

$$= -P_1(\delta = 1)E_1(\log \Lambda_\tau^{-1} \mid \delta = 1) - P_1(\delta = 0)E_1(\log \Lambda_\tau^{-1} \mid \delta = 0)$$

$$\geq -P_1(\delta = 1)\log E_1(\Lambda_\tau^{-1} \mid \delta = 1) - P_1(\delta = 0)\log E_1(\Lambda_\tau^{-1} \mid \delta = 0).$$

Since

$$P_0(\delta = i) = E_1\big(1_{(\delta=i)}\Lambda_\tau^{-1}\big)$$

$$= E_1\big(1_{(\delta=i)}E_1(\Lambda_\tau^{-1} \mid \delta = i)\big)$$

$$= P_1(\delta = i)E_1(\Lambda_\tau^{-1} \mid \delta = i), \qquad (1.36)$$

the inequality (1.35) becomes

$$
\begin{aligned}
E_1\Big(\frac{1}{2}\int_0^\tau \|\hat{h}_s\|^2\, ds\Big) &\geq -P_1(\delta = 1)\log\Big(\frac{P_0(\delta = 1)}{P_1(\delta = 1)}\Big) - P_1(\delta = 0)\log\Big(\frac{P_0(\delta = 0)}{P_1(\delta = 0)}\Big) \\
&= (1 - P_1(\delta = 0))\log\Big(\frac{1 - P_1(\delta = 0)}{P_0(\delta = 1)}\Big) - P_1(\delta = 0)\log\Big(\frac{1 - P_0(\delta = 1)}{P_1(\delta = 0)}\Big) \\
&= \log\Big(\frac{1 - P_1(\delta = 0)}{P_0(\delta = 1)}\Big) - P_1(\delta = 0)\log\Big(\frac{(1 - P_1(\delta = 0))(1 - P_0(\delta = 1))}{P_0(\delta = 1)P_1(\delta = 0)}\Big) \\
&\geq \log\Big(\frac{1 - \beta}{\alpha}\Big) - \beta\log\Big(\frac{(1 - \beta)(1 - \alpha)}{\alpha\beta}\Big) \\
&= (1 - \beta)\log\frac{1 - \beta}{\alpha} + \beta\log\frac{\beta}{1 - \alpha} \\
&= E_1\Big(\frac{1}{2}\int_0^{\tau^*}\|\hat{h}_s\|^2\, ds\Big). \qquad (1.37)
\end{aligned}
$$

The inequality

$$E_0\Big(\int_0^\tau \|\hat{h}_s\|^2\, ds\Big) \geq E_0\Big(\int_0^{\tau^*}\|\hat{h}_s\|^2\, ds\Big)$$

can be established in a similar fashion. $\qquad\square$

1.4 Bayesian Formulation

For the Bayesian formulation, let H be an $\{0, 1\}$-valued RV indicating the true hypothesis. By φ we denote the a priori probability that hypothesis H_1 is true. We consider a probability measure P on $(\Omega, \mathcal{F})$ such that

$$P(H = 1) = \varphi \qquad P(H = 0) = 1 - \varphi \qquad (1.38)$$

and such that for every $A \in \mathcal{F}$

$$P(A) = \varphi P_1(A) + (1 - \varphi) P_0(A), \qquad (1.39)$$

where P_1 and P_0 are the measures defined in Section 1.1.

We shall assume the cost of observation to accrue according to $k \int_0^t \|\hat{h}_s\|^2 ds$, where $k > 0$ and $\{\hat{h}_t, \ t \geq 0\}$ is defined by (1.8). The average cost due to data collection is simply

$$J_1(\tau) = E\left(k \int_0^\tau \|\hat{h}_t\|^2 dt \right) \qquad (1.40)$$

where the expectation is taken under P. The costs associated with the binary decision δ are given by

$$C(H, \delta) = \begin{cases} c_1, & \text{when } H = 1 \text{ and } \delta = 0; \\ c_2, & \text{when } H = 0 \text{ and } \delta = 1; \\ 0, & \text{otherwise.} \end{cases} \qquad (1.41)$$

where $c_1 > 0$ and $c_2 > 0$. The average cost due to the selection of δ is

$$\begin{aligned} J_2(\delta) &= E[\, C(H, \delta)\,] \\ &= c_1 P(H = 1, \delta = 0) + c_2 P(H = 0, \delta = 1) \\ &= c_1 \varphi P_1(\delta = 0) + c_2 (1 - \varphi) P_0(\delta = 1). \end{aligned} \qquad (1.42)$$

If $u = (\tau, \delta)$ is any admissible policy, then the corresponding average expected cost is

$$J(u) = E\left(k \int_0^\tau \|\hat{h}_s\|^2 ds + C(H, \delta) \right). \qquad (1.43)$$

The Bayesian approach to sequential detection requires the solution of the following optimization problem.

Problem $(\mathcal{P}_B)$: *Given* $\varphi \in (0, 1)$, *find* u^* *in* $\mathcal{U}$ *such that,*

$$J(u^*) = \inf_{u \in \mathcal{U}} J(u). \tag{1.44}$$

Let $\pi_t = P(H = 1 \mid \mathcal{F}_t^y)$ be the a posteriori probability of the hypothesis H_1 given $\mathcal{F}_t^y$. From the definition of π_t we see that

$$
\begin{aligned}
P(H = 1 \mid \mathcal{F}_t^y) &= E(H \mid \mathcal{F}_t^y) \\
&= \varphi E\left(\frac{dP_1}{dP} \mid \mathcal{F}_t^y \right) \\
&= \varphi E\left(\frac{dP_1}{d(\varphi P_1 + (1 - \varphi)P_0)} \mid \mathcal{F}_t^y \right) \\
&= \varphi \frac{dP_1}{d(\varphi P_1 + (1 - \varphi)P_0)}\left(\mathcal{F}_t^y \right)
\end{aligned}
\tag{1.45}
$$

From Girsanov's Theorem, we know that

$$\frac{dP_1}{dP_0}\left(\mathcal{F}_t^y \right) = \Lambda_t, \quad t \geq 0,$$

hence (1.45) becomes

$$\pi_t = P(H = 1 \mid \mathcal{F}_t^y) = \frac{\varphi \Lambda_t}{\varphi \Lambda_t + (1 - \varphi)}. \tag{1.46}$$

Lemma 1.4.1. *Let* $u = (\tau, \delta)$ *be any policy in* $\mathcal{U}$ *with* $\alpha(u) + \beta(u) < 1$ *and pose*

$$\alpha := \alpha(u) \qquad \beta := \beta(u).$$

If $u^* = (\tau^*, \delta^*)$ *is the threshold policy in* Γ *with thresholds* (A^*, B^*) *defined by*

$$A^* = \frac{\beta}{1 - \alpha}, \qquad B^* = \frac{1 - \beta}{\alpha},$$

then

$$J(u) = J_1(\tau) + J_2(\delta) \geq J_1(\tau^*) + J_2(\delta^*) = J(u^*). \tag{1.47}$$

Proof: This lemma follows from Corollary 1.2.3 and Theorem 1.3.1 since

$$J_2(\delta) = E(C(H,\delta)) = E(C(H,\delta^*)) = J_2(\delta^*).$$ □

For this problem, the requirement $\alpha(u) + \beta(u) < 1$ is not restrictive. In fact, for any u in $\mathcal{U}$ with $\alpha(u) + \beta(u) \geq 1$, the threshold policy $u^* = (\tau^*, \delta^*)$ in Γ, defined by $\tau^* = 0$ and

$$\delta^* = \begin{cases} 0, & c_1\varphi > c_2(1-\varphi) \\ 1, & c_1\varphi \leq c_2(1-\varphi) \end{cases}$$

incurs lower cost than u.

The main consequence of Lemma 1.4.2 is that now in problem $(\mathcal{P}_B)$, we only have to infimum over threshold policies. We now show the infimum in $(\mathcal{P}_B)$ is obtained by a threshold policy.

Theorem 1.4.1. *There exists a threshold policy u^* in Γ that solves problem $(\mathcal{P}_B)$. The optimal thresholds $0 < A^* \leq 1 \leq B^* < \infty$ with $A^* \neq B^*$, are given by the relations*

$$A^* = \left(\frac{1-\varphi}{\varphi}\right)\left(\frac{a^*}{1-a^*}\right), \qquad B^* = \left(\frac{1-\varphi}{\varphi}\right)\left(\frac{b^*}{1-b^*}\right). \tag{1.48}$$

where a^* and b^* are the unique solutions of the transcendental equations

$$c_2 + c_1 = k(\Psi'(a^*) - \Psi'(b^*)) \tag{1.49}$$

$$c_2(1-b^*) = c_1a^* + (b^* - a^*)(c_1 - k\Psi'(a^*)) + k(\Psi(b^*) - \Psi(a^*)), \tag{1.50}$$

with

$$\Psi(x) = (1-2x)\log\frac{x}{1-x}. \tag{1.51}$$

satisfying $0 < a^* < b^* < 1$.

Proof: It follows from Lemma 1.4.1 that in order to solve $(\mathcal{P}_B)$ we only need to consider threshold policies. Without loss in generality we assume $\tau^* > 0$. From

(1.24), (1.25) and (1.42) *the Bayesian cost for a threshold policy u is given by*

$$J(u) = J_1(\tau) + J_2(\delta)$$

$$= \varphi E_1(k \int_0^\tau \|\hat{h}_t\|^2 dt) + (1 - \varphi) E_0(k \int_0^\tau \|\hat{h}_t\|^2 dt) + c_1 \varphi \beta(u) + c_2 (1 - \varphi) \alpha(u)$$

$$= 2\varphi \, k \, w(\beta(u), \alpha(u)) + 2 (1 - \varphi) \, k \, w(\alpha(u), \beta(u)) + c_1 \varphi \, \beta(u) + c_2 (1 - \varphi) \, \alpha(u)$$

Therefore, the *optimal* threshold policy u^* is given by

$$J(u^*) = \inf_{u \in \Gamma} J(u) \tag{1.52}$$

$$= \inf_{\substack{\alpha, \beta \in (0,1) \\ \alpha + \beta < 1}} \left(2\varphi \, k \, w(\beta, \alpha) + 2 (1 - \varphi) \, k \, w(\alpha, \beta) + c_1 \varphi \, \beta + c_2 (1 - \varphi) \, \alpha \right).$$

After some algebraic simplification using equations (1.16), (1.48), and (1.51), equation (1.52) becomes

$$J(u^*) = \inf_{0 < a < b < 1} K(a, b) \tag{1.53}$$

with

$$K(a, b) \triangleq \left(2k\Psi(\varphi) + \left[(\varphi - a(c_2(1 - b) - 2k\Psi(b)) \right. \right.$$

$$\left. \left. + (b - \varphi)(c_1 a - 2k\Psi(a)) \right] / (b - a) \right).$$

The infimum in (1.53) is over an open set therefore, if the minimum value of $K(a, b)$ exists then it can be found by solving

$$\frac{\partial}{\partial a} K(a^*, b^*) = \frac{(b^* - \varphi)}{(b^* - a^*)^2} \left(2k [\Psi(b^*) - \Psi(a^*)] - 2k(b^* - a^*)\Psi'(a^*) \right.$$

$$\left. - c_2(1 - b^*) + c_1 b^* \right) = 0 \tag{1.54}$$

$$\frac{\partial}{\partial b} K(a^*, b^*) = \frac{(\varphi - a^*)}{(b^* - a^*)^2} \left(2k [\Psi(b^*) - \Psi(a^*)] - 2k(b^* - a^*)\Psi'(b^*) \right.$$

$$\left. - c_2(1 - a^*) + c_1 a^* \right) = 0 \tag{1.55}$$

This implies that

$$c_2 + c_1 = k(\Psi'(a^*) - \Psi'(b^*)), \tag{1.56}$$

$$c_2(1 - b^*) = c_1 a^* + (b^* - a^*)(c_1 - k\Psi'(a^*)) + k(\Psi(b^*) - \Psi(a^*)). \qquad (1.57)$$

It can be shown that for $0 < a^* < b^* < 1$, these equations have a unique solution (Shiryayev [1977, pp. 183–184]). Thus the thresholds are uniquely determined by c_1, c_2, and k. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

1.5 General Hypopaper Testing Problem

We will now *state* the results for the hypothesis testing problem when H_0 contains a signal in addition to noise. Here, $\{x_t^1,\ t \geq 0\}$ and $\{x_t^2,\ t \geq 0\}$ represent signal processes and are $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$-valued, respectively. The standard Brownian motions $\{w_t^1,\ t \geq 0\}$ and $\{w_t^2,\ t \geq 0\}$ are $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$-valued, respectively and are independent of the $\mathbb{R}^p$-valued standard Brownian motion $\{v_t,\ t \geq 0\}$. Under each hypothesis the observed data, dy_t, is the output of a stochastic differential equation

$$\text{Under } H_1: \qquad dy_t = h^1(x_t^1)\,dt + dv_t$$

$$dx_t^1 = f^1(x_t^1)\,dt + g^1(x_t^1)\,dw_t^1$$

$$\text{Under } H_0: \qquad dy_t = h^2(x_t^2)\,dt + dv_t$$

$$dx_t^2 = f^2(x_t^2)\,dt + g^2(x_t^2)\,dw_t^2$$

The functions f^1, f^2, g^1, g^2, h^1, and h^2 satisfy the Lipschitz and growth conditions of Section 1.1.

Let $\hat{h}_t^i = E_1(\,h^i(x) \mid \mathcal{F}_t^y\,)$, $i = 1,2$. In addition to (T1)-(T3) we assume that

$$P_i\left(\int_0^\infty \|\hat{h}_s^1 - \hat{h}_s^2\|^2\,ds = \infty \right) = 1, \quad i = 1,1.$$

Then, the solutions to the Bayesian and fixed probability of error formulations are still valid with Λ_t defined by

$$\Lambda_t = \exp\left(\int_0^t (\hat{h}_s^1 - \hat{h}_s^2)\,dy_s - \frac{1}{2} \int_0^t (\|\hat{h}_s^1\|^2 - \|\hat{h}_s^2\|^2)\,ds \right).$$

1.6 Summary

We have shown that in both formulations of the sequential hypothesis testing problem, the optimum decision policy is of threshold type. For the fixed probability of error formulation, this result was shown in (Liptser & Shiryayev [1978]). To our knowledge, no one has given explicit threshold formulas for the Bayesian case. Our ability to do so lies in our modification of the Bayesian cost. Usually, the Bayesian cost includes a constant penalty for each observation. We have chosen the observation cost to depend on the observation itself. This cost function is quite reasonable since it increases the penalty when the confidence increases and results in explicit formulas for the thresholds.

In order to implement the optimum policy, Λ_t needs to be computed from the observations $\{y_s, \ s \leq t\}$. In the next chapter we show that under the appropriate assumptions on the functions f, g, and h, the computation of Λ_t can be accomplished by calculating the unnormalized conditional density of x given the observations and then integrating.

2. Numerical Treatment

2.1 Introduction

In this chapter we will discuss the numerical method used to approximate Λ_t. It will be shown that $\Lambda_t = \int_{\mathbb{R}^n} u(x,t)\,dx$ where $u(x,t)$ is the solution to the *Zakai equation*. Our strategy will be to find a good approximation to $u(x,t)$ and then to can use it to approximate Λ_t. Therefore most of this chapter will be spent approximating $u(x,t)$. Since the *Zakai equation* is a *linear* stochastic partial differential equation, we will use standard numerical techniques for linear partial differential equations to obtain approximations to its solution. This means using semigroup methods to prove the necessary results and hence present one convergence theorem which can be used to check convergence of several approximation schemes.

This chapter is organized as follows: Section 2 contains the necessary results from nonlinear filtering. Sections 3-4 contain the necessary results from semigroups and parabolic partial differential equations. The remaining sections contain the approximation theorems for Λ_t and $u(x,t)$.

2.2 Results from Nonlinear Filtering

From the theory of nonlinear filtering, it is known (Liptser & Shiryayev [1977]) that under the appropriate conditions on f, g, and h, the unnormalized density of x given the observations satisfies the *linear* stochastic partial differential equation, known as

the *Zakai equation*, given below

$$
\begin{cases}
du(x,t) = L^* u(x,t)\, dt + u(x,t)\, h^T(x)\, dy_t \\[1mm]
u(x,0) = p_0(x) \\[1mm]
L^* u(x,t) = \displaystyle\sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\left[\sigma_{ij}(x)\, u(x,t)\right] - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left[f_i(x)\, u(x,t)\right] \\[1mm]
\sigma(x) = \dfrac{1}{2} g(x)\, g^T(x) \qquad (x,t) \in \Omega
\end{cases}
\qquad (2.1)
$$

where $p_0(x)$ is the initial density of x and $\Omega = \mathbb{R}^n \times [0,T]$. In order to guarantee existence and uniqueness of the solution of (2.1) (Baras, Blankenship & Hopkins, Jr [1983]), we make the following assumptions which are enforced throughout.

(A1) L^* is *uniformly elliptic*, that is, for some $\lambda > 0$, for all x, z in $\mathbb{R}^n$

$$
z^T \sigma(x)\, z \geq \lambda\, z^T z.
$$

(A2) The functions $f(x)$, $\sigma(x)$ and $h(x)$, along with $\frac{\partial}{\partial x_i} f_i(x)$, $\frac{\partial}{\partial x_i}\sigma_{ij}(x)$, $\frac{\partial}{\partial x_j}\sigma_{ij}(x)$, $\frac{\partial}{\partial x_i} h_k(x)$, $\frac{\partial^2}{\partial x_i \partial x_j} h_k(x)$, and $\frac{\partial^2}{\partial x_i \partial x_j}\sigma_{ij}(x)$ for $i,j = 1, \ldots, n$, and $k = 1, \ldots p$, are *uniformly bounded* and *Lipschitz continuous*.

In order to prove the necessary convergence results we will transform (2.1) into (2.7) (given below), approximate the solution of (2.7) and then transform back to get an approximation to the solution of (2.1). This transformation is discussed in (Clark [1978]) and is accomplished by a gauge transformation as follows.

Under assumption (A2) we see that

$$
L^* u = \sum_{i,j=1}^{n} \sigma_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} u + \sum_{j=1}^{n}\left(-f_j(x) + \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\sigma_{ij}(x)\right)\frac{\partial}{\partial x_j} u + c(x)\, u \quad (2.2)
$$

where

$$
c(x) = \sum_{j=1}^{n}\left(-\frac{\partial}{\partial x_j} f_j(x) + \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\sigma_{ij}(x)\right). \qquad (2.3)
$$

To simplify the notation let

$$
L^* u = A^* u + c(x)u, \qquad (2.4)
$$

and define

$$\varphi(x, y_t, t) = h^T(x)y_t - \frac{1}{2}\|h(x)\|^2 t + c(x)t. \tag{2.5}$$

Consider the function $r(x, t)$ which is related to the solution of (2.1) via

$$r(x, t) = e^{-\varphi(x, y_t, t)} u(x, t). \tag{2.6}$$

It follows from Itô's rule and the definition of the process $\{y_t,\ t \geq 0\}$ that

$$\begin{cases} \dfrac{\partial}{\partial t}r(x, t) = e^{-\varphi(x, y_t, t)} A^*[e^{\varphi(x, y_t, t)} r(x, t)] \\[2mm] r(x, 0) = p_0(x), \qquad (x, t) \in \Omega. \end{cases} \tag{2.7}$$

The importance of this transformation is that (2.7) is a classical parabolic partial differential equation with coefficients which (for every ω) are Hölder continuous functions of time. Therefore we can use standard techniques to prove convergence of approximations of (2.7) to its solution and by the simple transformation process in (2.6), get convergence of the respective approximation to the solution of (2.1). It is important to note that the only necessary term in the transformation is $h^T(x)\,y_t$. The other terms in the definition of φ help give desirable numerical properties to the subsequent approximation scheme.

Once the solution to (2.1) is found, Λ_t is calculated by integrating $u(x, t)$, i.e.,

$$\Lambda_t = \int_{\mathbb{R}^n} u(x, t)\, dx. \tag{2.8}$$

To see this note that from (1.7) and Itô's rule it follows that Λ_t satisfies the stochastic differential equation

$$\begin{cases} d\Lambda_t = \Lambda_t \hat{h}_t^T\, dy_t \\[2mm] \Lambda_0 = 1. \end{cases} \tag{2.9}$$

Let $< f, g >$ denote the L_2-inner product of f and g, i.e.,

$$< f, g > = \int_{\mathbb{R}^n} f(x)\, g(x)\, dx \tag{2.10}$$

then from (2.1)

$$d < u_t, 1 > = < L^* u_t, 1 > \ dt + < u_t \, h^T, 1 > \ dy_t$$

$$= < u_t, L \, 1 > \ dt + < u_t, h^T > \ dy_t$$

$$= \frac{< u_t, h^T >}{< u_t, 1 >} < u_t, 1 > \ dy_t$$

$$= < u_t, 1 > \hat{h}_t^T \ dy_t. \tag{2.11}$$

Therefore

$$\begin{cases} d < u_t, 1 > = < u_t, 1 > \hat{h}_t^T \ dy_t \\ \\ < u_0, 1 > = 1. \end{cases} \tag{2.12}$$

From existence and uniqueness of the solution of the stochastic differential equation (2.9) we obtain

$$\Lambda_t = < u_t, 1 > = \int_{\mathbb{R}^n} u(x, t) \, dt \qquad \text{a.s.}$$

Note that in the calculation above, we used the fact that $< u_t, 1 > \neq 0$. This follows from (A1)-(A2).

In practical applications, x_t and dy_t correspond to physical signals and as such, they are bounded. Therefore the boundness assumptions in (A2) on f, g, and h are not restrictive *since we can always take them bounded*. In fact because of the boundness of x_t, we need only consider values of $u(x, t)$ for x in a bounded set $D \subset \mathbb{R}^n$. Therefore, we will solve (2.1) with $\Omega = D \times [0, T]$ and with the boundary condition $u(x, t) = 0$ for x on ∂D. The assumptions (A1) and (A2) now hold with $\mathbb{R}^n$ replaced by D.

The boundness assumptions in (A2) on the partial derivatives of f, σ , and h are strong. We make these assumptions in order to prove that each suitable approximation, $u_n(x, t)$, converges uniformly to the true solution $u(x, t)$ of (2.1), i.e., for each t in $[0, T]$ and under the sup-norm

$$\lim_{n \to \infty} \| u_n(x, t) - u(x, t) \|_\infty = 0. \tag{2.13}$$

Since we will be implementing the resulting approximation scheme on a digital computer, we must be sure that under the finite arithmetic of the computer, Λ_t^n provides a good approximation to Λ_t (with the obvious notation). Therefore, we insist on uniform convergence. Notice that if (2.13) holds then by Hölder's inequality Λ_t^n converges to Λ_t. We note that under milder conditions on the functions f, g, and h, the corresponding weak convergence results can be shown (Kushner [1977]).

2.3 Definitions and Semigroup Results

Let X be a Banach space with norm $\|\cdot\|$ and let X_n be a Banach space with norm $\|\cdot\|_n$.

Definition 2.3.1. *The sequence of Banach spaces $\{X_n\}_0^\infty$ approximate X if there exist bounded linear operators $P_n : X \to X_n$, such that for all f in X,*

$$\lim_{n \to \infty} \|P_n f\|_n = \|f\|. \tag{2.15}$$

Note that it is not necessary for X_n to be a subset of X. For our purposes X_n will usually be finite dimensional while X will always be infinite dimensional.

Definition 2.3.2. *The sequence $\{f_n\}_0^\infty$, with f_n in X_n, converges to f in X, denoted $\lim_{n \to \infty} f_n = f$, if*

$$\lim_{n \to \infty} \|f_n - P_n f\|_n = 0. \tag{2.16}$$

Let $B(X, Y)$ denote *the set of all bounded linear operators from X into Y* and let $B(X) := B(X, X)$. The operator norm on $B(X)$ will also be denoted by $\|\cdot\|$ and the operator norm on $B(X_n)$ will be denoted $\|\cdot\|_n$.

For each $t \geq 0$, let $T(t)$ be a continuous linear map in X.

Definition 2.3.3. $T(t)$ *is said to be a strongly continuous semigroup, denoted (C_0)-semigroup, if it satisfies the following three properties:*

(A) $T(0) = I$

(B) $T(s + t) = T(s) \cdot T(t), \qquad s, t \geq 0$

(C) $\lim_{t \downarrow 0} T(t) f = f \qquad f \in X.$

It is known (Yosida [1980]) that for every (C_0)-semigroup $T(t)$ there exists constants $M \geq 1$ and $b \in \mathbb{R}$ such that

$$\|T(t)\| \leq M e^{bt} \qquad t \geq 0. \tag{2.17}$$

If $b = 0$ and $M = 1$ we say $T(t)$ is a *contraction* (C_0)-semigroup.

Definition 2.3.4 *The generator A of a (C_0)-semigroup is defined to be the unique linear operator A satisfying*

$$\lim_{t \downarrow 0} \left\| \frac{T(t)f - f}{t} - Af \right\| = 0, \qquad \text{for all } f \text{ in } D(A) \qquad (2.18)$$

where $D(A)$, the domain of A, is defined by all f in X where the limit exists.

Every (C_0)-semigroup has an infinitesimal generator, therefore we will use $\exp(tA)$ to denote a (C_0)-semigroup with infinitesimal generator A.

Lemma 2.3.5. *Let X be a Banach space. Suppose S and S^{-1} are bounded operators in X. Let $T(t)$ be a (C_0)-semigroup in X with infinitesimal generator A. Then $S^{-1}T(t)S$ is a (C_0)-semigroup in X with infinitesimal generator $S^{-1}AS$.*

Proof: First we show $S^{-1}T(t)S$ is a (C_0)-semigroup. The conditions (A) and (B) being clearly satisfied, we only need to show (C) is satisfied.

Let $g = Sf$ then

$$\left\| S^{-1}T(t)Sf - f \right\| \leq \left\| S^{-1} \right\| \left\| T(t)Sf - Sf \right\|$$

$$= \left\| S^{-1} \right\| \left\| T(t)g - g \right\| \to 0.$$

Therefore, $S^{-1}T(t)S$ is a (C_0)-semigroup, and we now show that $S^{-1}AS$ is its infinitesimal generator. For all f in X such that $Sf \in D(A)$

$$\lim_{t \downarrow 0} \left\| \frac{S^{-1}T(t)Sf - f}{t} - (S^{-1}AS)f \right\| = \lim_{t \downarrow 0} \left\| S^{-1} \left(\frac{T(t)Sf - Sf}{t} - ASf \right) \right\|$$

$$= \left\| S^{-1} \right\| \lim_{t \downarrow 0} \left\| \frac{T(t)g - g}{t} - Ag \right\|$$

$$= 0. \qquad \qquad \Box$$

Definition 2.3.6. *A linear operator A is closed if the graph of A defined by*

$$G(A) = \{(f, Af) : \quad f \in D(A)\} \qquad (2.19)$$

is a closed subspace of $X \times X$.

Definition 2.3.7. *A linear operator A in a Banach space X is closeable if the closure of the graph $\mathcal{G}(A)$ is itself the graph of a linear operator in X. The closure of A is denoted $\bar{A}$.*

Definition 2.3.8. *Let the sequence of Banach spaces $\{X_n\}_0^\infty$ approximate X. The linear operators A_n in X_n converge to the linear operator A in X, denoted $A = \lim_{n\to\infty} A_n$, if for all f in $\mathcal{D}(A)$*

$$\lim_{n\to\infty} \left\| A_n\, P_n\, f - P_n\, A\, f \right\|_n = 0. \tag{2.20}$$

The following theorem from Hille and Yosida characterizes the linear operators in X which generate (C_0)-semigroups.

Theorem 2.3.9 (Hille-Yosida). *Let A be a closed linear operator with domain and range in a Banach space X. Let $b \in \mathbb{R}$ and $M \geq 1$. The following assertions are equivalent:*

(i) The operator A generates a (C_0)-semigroup, $T(t)$ for which

$$\left\| T(t) \right\| \leq M\, e^{b\, t} \qquad t \geq 0. \tag{2.21}$$

(ii) The domain of A is dense in X, $sI - A$ is boundedly invertible for $s > b$ and

$$\left\| (s - b)^n\, (sI - A)^{-n} \right\| \leq M, \qquad n \in \mathbb{N} \tag{2.22}$$

Proof: See (Yosida [1980, pp. 246 –248]).

The following approximation theorem for (C_0)-semigroups comes from Trotter.

Theorem 2.3.10 (Trotter). *Let $\{h_n\}_0^\infty$ be a sequence of positive numbers converging to zero, and $\{T_n\}_0^\infty$ a sequence of linear operators in X_n satisfying the stability condition*

$$\left\| T_n^k \right\|_n \leq M\, e^{b\, k\, h_n} \tag{2.23}$$

where M and b are constants, independent of n and k. Let $A_n = h_n^{-1}(T_n - I)$ so that $T_n^k = (I + h_n A_n)^k$ and define $A := \lim_{n \to \infty} A_n$. If

  (i) $\mathcal{D}(A)$ is dense in X, and

  (ii) For some $s > b$, the range of $(sI - A)$ is dense in X,

then the closure of A is the infinitesimal generator of a (C_0)-semigroup $T(t)$ and

$$T(t) = \lim_{n \to \infty} (T_n)^{\left[\frac{t}{h_n}\right]}, \quad t \geq 0. \tag{2.24}$$

Proof: See (Trotter [1958, pp. 903–904]).

Since the operator $A(t)$ in (2.7) depends on time we are interested in approximating the solution of the abstract Cauchy problem for linear problems. In the Cauchy problem we want to find $u(t)$ in X such that

$$\begin{cases} \dfrac{du(t)}{dt} = A(t)\, u(t) \\ u(0) = u_0 \end{cases} \tag{2.26}$$

where for each $t \geq 0$, A(t) is a linear operator in a Banach space X. We assume that (2.26) is well-posed, i.e., there exists a unique solution to (2.26) in X.

If $A(t) \equiv A$ then the solution $u(t)$ to (2.26) is given by

$$u(t) = T(t)\, u_o \tag{2.27}$$

where A is the infinitesimal generator of the semigroup $T(t)$. In this case, Theorem 2.3.10 describes how to approximate $T(t)$ and hence how to approximate $u(t)$.

Definition 2.3.12. *The operator $U(t,s)$ is called a fundamental solution to (2.26) if it satisfies*

  (1) $U(t,s)$ is a strongly continuous function, defined for $0 \leq s \leq t \leq T$ which takes values in $B(X)$.

  (2) $U(t,r)\, U(r,s) = U(t,s)$ for $0 \leq s \leq r \leq t \leq T$.

(3) $U(s,s) = I$ *for all* $s \in [0,T]$.

(4) $\frac{\partial}{\partial t}U(t,s) = A(t)\,U(t,s)$.

(5) $\frac{\partial}{\partial s}U(t,s) = -U(t,s)A(s)$.

Conditions (4)–(5) are understood to hold on a dense subspace of X where they make sense. The derivatives $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial s}$ are taken in the strong topology of X. Notice that if $A(t) \equiv A$ then $U(t,s)$ is just the (C_0)-semigroup $T(t-s)$ with generator A. The solution to (2.26) is given by

$$u(t) = U(t,0)\,u_0. \tag{2.28}$$

where $U(t,s)$ is the fundamental solution to (2.26).

Let X be a Banach space with norm $\|\cdot\|$. Let Y be a dense subspace of X and assume Y is itself a Banach space, with norm $\|\cdot\|_Y$. Suppose there exists a constant c such that $\|v\| \leq c\|v\|_Y$ for all v in Y. Henceforth, we assume X and Y satisfy these conditions. We now give some useful definitions.

Definition 2.3.13. *The set of all generators of (C_0)-semigroups with constants M and b in X is denoted by by $G(X,M,b)$, and the set of all generators of (C_0)-semigroups in X is denoted by*

$$G(X) := \bigcup_{b \in \mathbb{R}} \bigcup_{M \geq 1} G(X,M,b) \tag{2.29}$$

Definition 2.3.14. *Let $A \in G(X)$. The Banach space Y is called A-admissible if $\exp(t\,A)$ maps Y into Y, and if the restriction of $\exp(t\,A)$ to Y forms a semigroup in Y.*

Definition 2.3.15. *The family of operators $A(t)$ in $G(X)$ is called stable, with stability constants M and b, if there exist real numbers $M \geq 1$ and b such that*

$$\left\| \prod_{j=1}^{k} \exp\big(s_j\,A(t_j)\big) \right\| \leq M\,e^{b\,(s_1+\cdots+s_k)}, \qquad s_j \geq 0, \tag{2.30}$$

for all $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k \leq T$ *and* $k = 1, 2, \ldots$. *Here the product is time ordered, i.e.*

$$\prod_{j=1}^{k} \exp\big(s_j \, A(t_j) \big) = \exp\big(s_k \, A(t_k) \big) \cdots \exp\big(s_1 \, A(t_1) \big). \qquad (2.31)$$

For the remainder of this chapter, all operator products will be time ordered. The following lemma will be useful to test for stability.

Lemma 2.3.16. *Assume that* $A(t)$ *is stable with stability constants* M *and* b. *If for each* t *in* $[0, T]$, $B(t)$ *is bounded, i.e.,* $\|B(t)\| \leq K < \infty$ *then for each* t *in* $[0, t]$, $A(t) + B(t)$ *belongs to* $G(X)$ *and is stable with stability constants* M *and* $b + MK$.

Proof: See (Kato [1970, p 248]).

We now state assumptions on the generator $A(t)$ in (2.26).

(A3) $A(t)$ is stable with constants 1 and b.

(A4) The Banach space Y is A-admissible and $Y \subset \mathcal{D}(A(t))$ for each t in $[0, T]$ and the operator $A(t)$ is a continuous function in the norm of $B(Y, X)$.

(A5) $A(t)$ is closed in X.

(A6) For some $s > b$, the range of $(sI - A(t))$ is dense in X, where b is independent of t.

The following convergence theorem combines the theorems (Kato [1970, Theorem 4.1 p247]) and (Trotter [1958, Theorem 5.3 p903]).

Theorem 2.3.17. *Let* h_n *be a sequence of positive numbers such that* $h_n \to 0$ *as* $n \to \infty$. *Suppose that for each* t *in* $[0, T]$ *the linear operators* $T_{n,t}$ *in* X_n *satisfy the stability conditions*

$$\left\| T_{n,t}^{k} \right\|_{n} \leq M \, e^{k \, b \, h_n} \qquad (2.32a)$$

and

$$\left\| \prod_{j=0}^{N} T_{n,t_j} \right\|_{n} \leq M \, e^{b \, N \, h_n} \qquad (2.32b)$$

where $t_j = j\, h_n$, $t_N \leq T$ and the constants $M \geq 1$, $b \in \mathbb{R}$ are independent of t, k and n.

Let $A_n(t) = (T_{n,t} - I)/h_n$ and suppose that

$$A(t) := \lim_{n \to \infty} A_n(t). \tag{2.33}$$

for each t_k in $[0, T]$. For each pair (t, s) satisfying $0 \leq s < t \leq T$, $t_k \leq s < t_{k+1}$, and $t_l \leq t < t_{l+1}$, define the operator

$$U_n(t, s) = \begin{cases} I & k = l, \\ \displaystyle\prod_{j=k}^{l-1} T_{n,t_j} & k < l. \end{cases} \tag{2.34}$$

If the operator $A(t)$ in (2.33) satisfies (A3)–(A6) and if the corresponding Cauchy problem in well-posed then

$$U(t, s) = \lim_{n \to \infty} U_n(t, s) \tag{2.35}$$

where $U(t, s)$ is the fundamental solution generated by $A(t)$.

Proof: It is easy to see from its definition that $U_n(t, s)$ satisfies (1), (2), and (3) of Definition 2.3.12. Furthermore it can be seen that $U_n(t, s)$ maps X_n into X_n, and that for each v in Y,

$$\frac{U_n(t + h_n, s) - U_n(t, s)}{h_n} P_n v = A_n(t)\, U_n(t, s) P_n v \tag{2.37}$$

and

$$\frac{U_n(t, s + h_n) - U(t, s)}{h_n} P_n v = -U_n(t, s + h_n)\, A_n(s) P_n v. \tag{2.38}$$

From (2.32) and the stability of $A(t)$ we have that

$$\left\| U_n(t, s) \right\| \leq M\, e^{b(t-s)}, \qquad \left\| U(t, s) \right\|_Y \leq \tilde{M}\, e^{\tilde{b}(t-s)}. \tag{2.39}$$

Next $v \in Y$ and let

$$\bar{U}(t, s) = U\left(h_n\, [t/h_n],\, h_n\, [s/h_n] \right).$$

It follows that all we need to show is the convergence of $U_n(t,s)$ to $\bar{U}(t,s)$ hence

$$
\begin{aligned}
\left\| U_n(t,s)P_n v - P_n \bar{U}(t,s)v \right\|_n &= \left\| \sum_{j=k}^{l-1} U_n(t,t_{j+1})P_n \bar{U}(t_{j+1},s)v - U_n(t,t_j)P_n \bar{U}(t_j,s)v \right\|_n \\
&= \left\| \sum_{j=k}^{l-1} U_n(t,t_{j+1})P_n \bar{U}(t_{j+1},s)v - U_n(t,t_{j+1})P_n \bar{U}(t_j,s)v \right. \\
&\qquad \left. + U_n(t,t_{j+1})P_n \bar{U}(t_j,s)v - U_n(t,t_j)P_n \bar{U}(t_j,s)v \right\|_n \\
&= \left\| \sum_{j=k}^{l-1} U_n(t,t_{j+1})P_n (\bar{U}(t_{j+1},s) - \bar{U}(t_j,s))v \right. \\
&\qquad \left. + (U_n(t,t_{j+1}) - U_n(t,t_j))P_n \bar{U}(t_j,s)v \right\|_n \\
&= \left\| \sum_{j=k}^{l-1} U_n(t,t_{j+1})P_n (U(t_{j+1},t_j) - I)\bar{U}(t_j,s)v \right. \\
&\qquad \left. - U_n(t,t_{j+1})h_n A_n(t_j)P_n \bar{U}(t_j,s)v \right\|_n \\
&\leq M' e^{-\gamma(t-s)} \sum_{j=k}^{l-1} \left\| (U(t_{j+1},t_j) - I)v - h_n A_n(t_j)P_n v \right\|_n \\
&\leq M'(t-s)e^{-\gamma(t-s)} \\
&\qquad \sup_{j=k,\ldots,l-1} \left\| \frac{U(t_{j+1},t_j) - I}{h_n}v - A_n(t_j)P_n v \right\|_n \qquad (2.40)
\end{aligned}
$$

where $\gamma = \max\{b, \tilde{b}\}$ and $M' = M\tilde{M}$. From (2.33) and Definition 2.3.12 it follows that $U_n(t,s)P_n v$ converges as $n \to \infty$ uniformly in $0 \leq s \leq t \leq T$. From the fact that Y is dense in X it follows that for all u in X,

$$
\lim_{n \to \infty} \left\| U_n(t,s)P_n u - P_n U(t,s)u \right\| = 0 \qquad (2.41)
$$

exists uniformly in $0 \leq s \leq t \leq T$. $\square$

The results above can be used to show convergence of several different types of approximation schemes. In the next sections, we will apply them to implicit Euler type, finite difference approximations for the parabolic equation (2.7).

2.4 Results on Parabolic Partial Differential Equations

In this section we will quote results from the theory of parabolic partial differential equations to establish that A^* in (2.4) generates a (C_0)-semigroup and to show that there exists a unique solution to (2.7). This will done for the case where x in (2.7) is a scalar diffusion. Similar techniques can be applied to the vector case and results in schemes similar to those found in (Richtmyer & Morton [1967]).

Throughout this section, let $D = (a, b)$ a bounded interval and let $X = C_b[(a, b)] \cap C_0[(a, b)]$ under the sup-norm and $Y = C^\infty[(a, b)] \cap C_0[(a, b)]$ with norm $\|\cdot\|_Y$ such that for each f in Y

$$\|f\|_Y = \sum_{n=0}^{\infty} \|f^{(n)}\|.$$

It is known that Y forms a dense subspace in X. Furthermore, for every f in Y, $\|f\| \leq \|f\|_Y$. We will consider the Cauchy problem:

$$\begin{cases} u_t(x, t) = a(x, t)\, u_{xx}(x, t) + b(x, t)\, u_x(x, t) + c(x, t)\, u(x, t), \\ \qquad u(a, t) = u(b, t) = 0, \quad t \in [0, T] \\ u(x, 0) = u_0(x), \qquad (x, t) \in \Omega, \end{cases} \tag{2.43}$$

where $\Omega = (a, b) \times [0, T]$. The follow theorem from Besala gives the necessary existence result for the solution of (2.43).

Theorem 2.4.1 *(Besala) For (2.43), let the functions $a(x, t)$, $b(x, t)$, and $c(x, t)$ (real valued) together with $a_x(x, t)$, $a_{xx}(x, t)$, $b_x(x, t)$ be locally Hölder continuous in Ω. Assume that for all (x, t) in Ω*

   i) $a(x, t) \geq \lambda > 0$, for some λ

   ii) $c(x, t) \leq 0$,

   iii) $c(x, t) - b_x(x, t) + a_{xx}(x, t) \leq 0$.

Then the Cauchy problem (2.43) has a fundamental solution $U(t, s)$ in the Banach space X where

$$\|U(t, s)\| \leq 1. \tag{2.44}$$

Furthermore, there exists a function $\Gamma(x, t; z, s)$ which satisfies

$$0 \leq \Gamma(x, t; z, s) \leq \frac{k}{\sqrt{(t-s)}} \tag{2.45}$$

for some positive k, and

$$\int_a^b \Gamma(x, t; z, s)\, dz \leq 1$$

$$\int_a^b \Gamma(x, t; z, s)\, dx \leq 1 \tag{2.46}$$

such that

$$U(t, s)\, f(x) = \int_a^b \Gamma(x, t; z, s)\, f(z)\, dz.$$

Moreover, if $u_0(x)$ is continuous and bounded, then

$$u(x, t) = U(t, 0)u_o(x) = \int_a^b \Gamma(x, t; z, 0)\, u_0(z)\, dz \tag{2.47}$$

is a bounded solution of (2.43).

Proof: See (Besala [1975]).

The following theorem due to Friedman gives the necessary uniqueness result for (2.43).

Theorem 2.4.2. If $a(x, t)$, $b(x, t)$, and $c(x, t)$, in (2.43) are continuous in Ω. If there exists $\lambda > 0$ such that for all x in (a, b)

$$a(x, t) \geq \lambda \tag{2.48}$$

for (x, t) in Ω, then there exists at most one solution to the Cauchy problem (2.43).

Proof: See (Friedman [1964, Theorem 7, p. 41])

Corollary 2.4.3. For each t_k in $[0, T]$, the operator $A(t_k)$ defined in (2.7) generates a (C_0)-semigroup in X with constants 1 and K.

Proof: Using the notation of (2.43) for the operator $A(t_k)$ and from assumption (A2) it follows that there exists a finite constant $K > 0$ such that

$$\max_{x \in \mathbb{R}}\{a_{xx}(x,t) - b_x(x,t) - K\} \le 0. \tag{2.50}$$

Define $\bar{A}(t_k) = A(t_k) - K$. From assumption (A1) it follows that $\bar{A}(t_k)$ satisfies the hypotheses of Theorem 2.4.1, therefore $\bar{A}(t_k)$ generates a *contraction* (C_0)-semigroup. From Lemma 2.3.16 it follows that that $A(t_k)$ generates a (C_0)-semigroup in X with constants 1 and K.

Corollary 2.4.4. *Under assumptions (A1)-(A2) the semigroup generated by $A(t_k)$ for each t_k in $[0,T]$ is given by*

$$\exp\big(t\,A(t_k)\big) = e^{-\varphi(x,y_{t_k},t_k)} \exp\big(t\,A^*\big)\, e^{\varphi(x,y_{t_k},t_k)} \tag{2.51}$$

where $\varphi(\cdot,\cdot,\cdot)$ is defined in (2.23).

Proof: Directly from Lemma 2.3.5.

We are now ready to show a convergence theorem for the case when only time is discretized

Theorem 2.4.5. *Let $\Delta t = T/n$ and $k = [t/\Delta t]$. For $k > 0$ define*

$$U_n(t,0)p_0(x) = e^{-\varphi(x,y_{k\Delta t},k\Delta t)}\left(\prod_{j=0}^{k-1} e^{h(x)\,\Delta y_j + c(x)\Delta t - \frac{1}{2}h^2(x)\Delta t}\exp\big(\Delta t\,A^*\big)\right)p_0(x) \tag{2.52}$$

where $\Delta y_j = y_{(j+1)\Delta t} - y_{j\Delta t}$, then

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\left\|U_n(t,0)p_0(x) - r(x,t)\right\|_\infty = 0, \tag{2.53}$$

where $r(x,t)$ is the solution of (2.7). Furthermore, in view of (2.6), it follows that

$$\lim_{n\to\infty}\sup_{t\in[0,T]}\left\|e^{\varphi(x,y_{k\Delta t},k\Delta t)}U_n(t,0)p_0(x) - u(x,t)\right\|_\infty = 0 \tag{2.54}$$

where $u(x,t)$ is the solution of (2.1).

Proof: Let $T_{n,t} = \exp\big(\Delta t\, A(t)\big)$ where $A(t)$ is the operator in (2.7). Substituting (2.51) into the definition of $\tilde{U}_n(t,0)$ of Corollary 2.3.18 and rewriting the product yields (2.52). From Corollary 2.3.4, $A(t)$ is stable with constants 1 and K hence all of the hypothesis of Theorem 2.3.17 are satisfied, therefore the convergence (2.53).

Corollary 2.4.6. *Let $\tilde{A}_n(t) := A(t_{k+1})$ for $t_k \leq t < t_{k+1}$ replace $A_n(t)$ in Theorem 2.3.17. If $t_k = k\,\Delta t$ then for $A(t)$ in (2.7) and using (2.6), we see that*

$$\tilde{V}_n(t,0)p_0(x) = \left(\prod_{j=0}^{k-1} \exp\big(\Delta t\, A^*\big)\, e^{h(x)\,\Delta y_j + c(x)\Delta t - \frac{1}{2}h^2(x)\,\Delta t} \right) p_0(x) \qquad (2.55)$$

converges to $u(x,t)$ in (2.1), i.e.,

$$\lim_{n \to \infty} \sup_{t \in [0,T]} \left\| \tilde{V}_n(t,0)p_0(x) - u(x,t) \right\|_\infty = 0. \qquad (2.56)$$

Proof: All of the hypotheses of Theorem 2.3.17 are satisfied with $\tilde{A}_n(t)$, hence the result (2.35) with only the indices changed.

Example 1. Time discretization of (2.1).

Under the assumptions (A1)–(A2) the discrete time approximation scheme

$$v_n(x,t) = \left(\prod_{j=0}^{[\frac{t}{h_n}]} (I - \Delta t L^*)^{-1} e^{h(x)\,\Delta y_j - \frac{1}{2}h^2(x)\,\Delta t} \right) p_0(x) \qquad (2.57)$$

converges to the solution of (2.1).

Let $T_n = (I - \Delta t L^*)^{-1}$ and the Banach space X as above. It is known (Richtmyer & Morton [1967]) that

$$\|T_n\| \leq e^{K\Delta t} \qquad (2.58)$$

therefore

$$\exp\big(t\,L^*\big) = \lim_{n \to \infty} (T_n)^{[\frac{t}{\Delta t}]}. \qquad (2.59)$$

Convergence of (2.57) to the solution of (2.1) follows from Theorem 2.3.17 and Corollary 2.5.6. This scheme appears in (LeGland [1981]; Pardoux & Talay [1983]) and was obtained by different methods.

By arguments similar to those in Theorem 2.4.5 it is possible to show that

$$V_n(t,0)p(x) = \left(\prod_{j=0}^{k-1} e^{h^T(x)\Delta y_j - \frac{1}{2}\|h(x)\|^2 \Delta t} \exp\left(\Delta t \, L^* \right) \right) p_0(x) \qquad (2.60)$$

converges to the solution of (2.1), where Δy_j and Δt are defined in Theorem 2.4.5. This result is *important for applications* since it demonstrates the importance of the transition density of x in the calculation of $u(x,t)$. The importance of the transition density has been shown and is illustrated in the Kallianpur-Striebel formula. For us, it is interesting that (2.60) can be thought of as a direct approximation the Kallianpur-Striebel formula.

In applications, the transition density of x can usually be approximated from the observed data. While the functions f and g usually cannot. Therefore, this approximation method has the additional, practical advantage of just needing the transition density.

2.5 The Finite Difference Approximation Scheme for (2.1)

It is now possible to find several finite difference approximation schemes for (2.1). In fact, from Theorem 2.4.5 it follows that all we need is a good approximation for $\exp(\Delta t A^*)$. Several good approximation schemes exist in the numerical analysis literature, however we will only concentrate on one.

In the scalar x case we have $D = (a, b)$,

$$
\begin{aligned}
L^* u(x, t) &= a(x)\, u_{xx}(x, t) + b(x)\, u_x(x, t) + c(x) u(x, t) \\
&= A^* u(x, t) + c(x) u(x, t)
\end{aligned}
\tag{2.60}
$$

where

$$
\begin{aligned}
a(x) &= \frac{1}{2} g^2(x) \\
b(x) &= g(x)\, g'(x) - f(x) \\
c(x) &= g(x)\, g''(x) + (g'(x))^2 + g(x)\, g'(x) - f'(x)
\end{aligned}
\tag{2.61}
$$

Let $\Delta x > 0$ and define $x_k = a + k\,\Delta x$ and n such that $x_n \le b$. Consider the collection of points $\{x_k\}_0^n$ in D. Let $\Delta x \to 0$ as $n \to \infty$ such that $x_n \to b$ as $n \to \infty$. Let X and Y be the Banach spaces in Section 2.4 and let the approximating Banach spaces $X_n = \mathbb{R}^{n+1}$ under the ∞-norm. Define the operator $P_n : X \to X_n$ such that for each ϕ in X

$$
\left(P_n \phi\right)_i = \phi(x_i), \quad i = 0, \ldots, n.
\tag{2.62}
$$

It is easily checked that for each ϕ in X

$$
\lim_{n \to \infty} \left\| P_n \phi \right\|_n = \left\| \phi \right\|.
$$

The linear operator A_n in X_n will be obtained from A^* by replacing the x-derivatives in A^* with finite difference approximations. To this end, for each ϕ in Y pose

$$
\begin{aligned}
\left(A_n\, P_n\, \phi\right)_i = a(x_i) \frac{\phi(x_{i+1}) - 2\phi(x_i) + \phi(x_{i-1})}{(\Delta x)^2} &+ \max(b(x_i), 0) \frac{\phi(x_{i+1}) - \phi(x_i)}{\Delta x} \\
&+ \min(b(x_i), 0) \frac{\phi(x_i) - \phi(x_{i-1})}{\Delta x}
\end{aligned}
\tag{2.63}
$$

where i=0,...,n.

From Taylor's Theorem

$$\phi(x_i \pm \Delta x) = \phi(x_i) + \phi_x(x_i)(\pm \Delta x) + \frac{1}{2}\phi_{xx}(x_i + \theta_\pm)(\Delta x)^2 \qquad (2.64)$$

where $0 < \theta_+ < \Delta x$ and $-\Delta x < \theta_- < 0$. Substituting into (2.63) yields

$$(A_n P_n \phi)_i = \frac{a(x_i)}{2}\left(\phi_{xx}(x_i + \theta_+) + \phi_{xx}(x_i + \theta_-)\right) + b(x_i)\left(\phi_x(x_i) + O(\Delta x)\right). \qquad (2.65)$$

Since ϕ_{xx} is continuous it follows that

$$\lim_{n \to \infty} \left\| A_n P_n \phi - P_n A^* \phi \right\|_n = 0. \qquad (2.66)$$

We are now able to show convergence of the full discretization of (2.7), i.e. when both time and space are discretized.

Theorem 2.5.1. *Let $\Delta t = h_n$ and let P_n as in (2.62). Define*

$$\begin{aligned} D_k &= \text{diag } \{P_n\, e^{h^T(x)\, \Delta y_k + c(x)\Delta t - \frac{1}{2}h^2(x)\Delta t}\} \\ &= \text{diag } \{e^{h^T(x_i)\, \Delta y_k + (c(x_i) - \frac{1}{2}h^2(x_i))\, \Delta t}\} \end{aligned} \qquad (2.67)$$

where $\Delta y_k = y_{(k+1)\Delta t} - y_{k\Delta t}$. Consider v^k in $\mathbb{R}^{n+1}$ defined by

$$\begin{cases} v^{k+1} = (I - \Delta t\, A_n)^{-1} D_k v^k \\ v^0 = P_n p_0. \end{cases} \qquad (2.68)$$

Let $k\Delta t \to t$ as $n \to \infty$, then for every t in $[0, T]$

$$\lim_{n \to \infty} \sup_{i=0,\ldots,n} \left| (v^k)_i - u(x_i, k\Delta t) \right| = 0. \qquad (2.69)$$

Proof: Let $T_n = (I - \Delta t A_n)^{-1}$. From (2.63) we see that A_n is a diagonally dominant matrix. It has the form

$$A_n = \begin{pmatrix} - & + & & & \\ + & \ddots & \ddots & & \\ & \ddots & \cdot & + & \\ & & + & - \end{pmatrix}. \qquad (2.70)$$

For every $\Delta t > 0$, $(T_n)^{-1}$ has the form

$$(T_n)^{-1} = I - \Delta t A_n = \begin{pmatrix} + & - & & & \\ - & \ddots & \ddots & & \\ & \ddots & & - & \\ & & - & + \end{pmatrix}. \tag{2.71}$$

and is strictly diagonally dominant since the diagonal is reinforced. It is easily checked that $(T_n)^{-1}$ is an *inverse positive* matrix, i.e., $(T_n)_{ij} \geq 0$, see (Schröder [1978, Corollary 1.6b, p221]). Since $(T_n)^{-1}$ is invertible for all Δt in $[0, T]$ we see that

$$\|T_n\|_\infty \leq e^{b\Delta t}$$

for some b in $\mathbb{R}$, so that T_n satisfies (2.23).

Let $w^k = (E_k)^{-1} v^k$ where

$$\begin{aligned} E_k &= \mathrm{diag}\,\{P_n e^{\varphi(x, y_{k\Delta t}, k\Delta t)}\} \\ &= \mathrm{diag}\,\{e^{\varphi(x_i, y_{k\Delta t}, k\Delta t)}\} \end{aligned} \tag{2.72}$$

and $\varphi(\cdot, \cdot, \cdot)$ is defined in (2.5). Then from (2.68) we obtain

$$\begin{cases} w^{k+1} = (E_{k+1})^{-1} T_n E_{k+1} w^k \\ w^0 = P_n p_o \end{cases}. \tag{2.73}$$

Define $\widetilde{T}_n(t_k) = (E_{k+1})^{-1} T_n E_{k+1}$ where $t_k = k\Delta t$ then for each t_k in $[0, T]$

$$\|\widetilde{T}_n^k(t_k)\| \leq M e^{kb\Delta t}$$

and

$$\left\|\prod_{j=1}^{N} \widetilde{T}_n(t_j)\right\| \leq M e^{bN\Delta t}.$$

As in Theorem 2.3.10, we define

$$\begin{aligned} \tilde{A}_n(t_k) &= \frac{\widetilde{T}_n(t_k) - I}{\Delta t} \\ &= (E_{k+1})^{-1} \left(\frac{T_n - I}{\Delta t}\right) E_{k+1}. \end{aligned} \tag{2.74}$$

From the continuity and differentiability of $\varphi(\cdot,\cdot,\cdot)$ and from (2.66) it follows that for each ϕ in $\mathcal{D}(A^*)$

$$\lim_{n\to\infty} \left\| P_n\, e^{-\varphi(x,y_{t_k},t_k)} A^* e^{\varphi(x,y_{t_k},t_k)}\, \phi(x) - \tilde{A}_n(t_k)\, P_n\, \phi(x) \right\|_\infty = 0.$$

Therefore from Theorem 2.3.17, w^k in (2.73) converges to the solution of (2.7). The continuity of the transformation (2.6) implies (2.68) converges to the solution of (2.1) hence (2.69). □

The approximation for $b(x)\, u(x,t)$ is standard in the numerical literature and is motivated from stability of finite difference methods for hyperbolic equations (Richtmyer & Morton [1967, p. 292]). Basically, when $b(x_i) > 0$ we take a forward difference approximation and when $b(x_i) \leq 0$ we take a backward difference approximation. Using this approximation ensures positivity of our approximation scheme, independent of the value of Δt.

2.6 Convergence of the Corresponding Likelihood Ratio Approximation

Using v^k defined in (2.69) it is easy to construct a convergent approximation to the likelihood ratio. Let $\{x_k\}_0^n$ be the collection of points defined in Section 2.5 and let

$$u_n(x,t) = (v^k)_i \qquad \text{if} \quad x_i \leq x < x_{i+1}, \quad \Delta t \leq t < (k+1)\Delta t$$

and define

$$\Lambda_t^n = \int_a^b u_n(x,t)\, dx = \sum_{i=0}^n (v^k)_i\, \Delta x$$

then

$$\begin{aligned}
\lim_{n \to \infty} |\Lambda_t - \Lambda_t^n| &= \lim_{n \to \infty} \int_a^b u(x,t)\, dx - \int_a^b u_n(x,t)| \\
&\leq (b-a) \lim_{n \to \infty} \sup_{x \in (a,b)} |u(x,t) - u_n(x,t)| \\
&= (b-a) \lim_{n \to \infty} \sup_{i=0,\ldots,n} |u(x_i,t) - u_n(x_i,t)| \\
&= (b-a) \lim_{n \to \infty} \sup_{i=0,\ldots,n} |u(x_i,t) - (v^k)_i| \\
&= 0.
\end{aligned}$$

Therefore, Λ_t^n is a convergent approximation to Λ_t.

2.7 Summary

The main result in chapter was the convergence result in Theorem 2.3.17. This theorem presented verifiable conditions which ensure convergence of an approximation scheme for the abstract Cauchy problem (2.26). These conditions were used to show convergence of the natural finite difference approximation scheme for the solution to (2.1). This, in turn, implied show convergence of the approximate likelihood ratio to the true likelihood ratio.

Schemes similar to (2.68) have been discussed in (Kushner [1977]; Pardoux & Talay [1983]). Furthermore, numerical studies have been performed in (Yavin [1985]) using these methods which have produced satisfactory results for approximations to $\hat{h}_t$.

Again, the importance of the result in Theorem 2.4.5 is that the stochastic part of (2.1) has been isolated from the non-stochastic part. Therefore, the wealth of information on numerical approximation for parabolic partial differential equations can be used to approximate $\exp(t A^*)$, and this in turn is used to approximate (2.1).

3. VLSI Architectures

3.1 Introduction

In this chapter we will present a VLSI architecture for solving (2.1) when x is scalar. This architecture will *not* be efficient for vector x. However, as was pointed out in Section 2.9, it is possible to use some of the advanced techniques for sparse matrices to develop architectures for higher state dimensions. It is important to emphasize that these methods are not sufficiently efficient for real time implementations of problems with state dimensions higher than three.

In Section 2.7 we presented a finite difference scheme to approximate the solution of (2.1). At each time $t = k\Delta t$, this scheme involved solving the linear equation

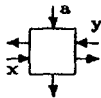$$(I - \Delta t A_n) V^{k+1} = D_k V^k, \tag{3.1}$$

where $D_k = \text{diag}\{\exp(h(x_i)\,\Delta y_k + (c(x_i) - \frac{1}{2}h^2(x_i))\,\Delta t)\}$. Our goal is to give a design for a VLSI chip to efficiently solve (3.1). This means that:

(1) the time necessary to compute V^{k+1}, given V^k, A, and y_k, should be below a problem dependent threshold; and

(2) the control structure within the chip should be simple and regular.

We will show how the *systolic array architecture* of (Kung & Leiserson [1980]) can be used to obtain a VLSI design satisfying these goals. The discussion of systolic architectures follows closely (Kung & Leiserson [1980]).

3.2 Systolic Arrays and Sequential Computations

Systolic processors are arrays of tightly synchronized simple processors in which data is fed to each processor in a regular and ordered manner. In these arrays, data flows through the processors much in the same way that blood flows through the heart, hence the term systolic. Because of their order and regularity, systolic processors are significantly faster than conventional processors.

We will be interested in systolic processors which perform linear algebra operations. The basic component of these arrays is the *inner product processor* (IPP) denoted by ⊣▭⊢ . At each clock pulse the IPP takes the inputs x, y and a and computes $ax + y$ (the inner product step). This value is output on the y-output line, and the x and a values pass through to their respective output lines untouched.

To illustrate the speedup, consider how a typical von Neumann computer would perform the inner product step. Before a von Neumann computer can execute an instruction it must first get (or fetch) the instruction from memory and then fetch the arguments (or operands) for that instruction from memory. Table 3.2.1 shows the computer instructions necessary to execute the inner product step. It does not include the overhead of fetching the instructions and the operands.

Instruction	(cont.)
1. Fetch x	3. Fetch y
1. Fetch a	5. Add (ax) & y
2. Multiply a & x	6. return $(ax + y)$

Table 3.2.1. *Operations performed executing the inner product step.*

However, the computer actually performs two additional operations in executing each ᵢmand. For each instruction listed in Table 3.2.1 ᵢ. ᵢ are three additional instructions to perform. Table 3.2.2 shows these additional instructions.

Therefore, performing the inner product step requires a von Neumann computer

Actual Operations
Fetch Instruction
Fetch Operand for Instruction
Execute Instruction

Table 3.2.2. *Operations performed each instruction in Table 3.1.*

to execute 18 operations. Notice that *only* two of these operations are mathematical, the other 16 involve data manipulation and program control. The time required to perform each of the 18 operations varies from computer to computer, however on any one computer, these steps take approximately the same time to perform (when using fixed point arithmatic). Systolic processors are fast because they eliminate the 16 wasted operations.

To illustrate how a systolic array can be used in a linear algebra operation we consider the matrix-vector operation $y = Ax$ where A is an n-by-n matrix and x is a vector in $\mathbb{R}^n$. Let y_i denote the ith component of the vector y. The following n-step recursion can be used to calculate y_i

$$y_i^{(1)} = 0$$
$$y_i^{(k+1)} = y_i^{(k)} + a_{ik} x_k, \quad 1 \le k \le n.$$

It is easily verified that $y_i = y_i^{(n+1)}$.

The above recursion shows that matrix-vector multiplication is nothing more than a sum of inner product steps. Therefore, it is reasonable to expect that with the correct data flow we can accomplish this matrix-vector multiplication with an array of IPP's.

Let A be an n-by-n matrix. We say A is a band matrix with bandwidth w if $a_{ij} = 0$ outside a w-wide diagonal band containing the main diagonal. For example,

the matrix A in (3.2) has bandwidth 3.

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & a_{2,3} & 0 & 0 \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & 0 \\ 0 & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} \\ 0 & 0 & a_{53} & a_{54} & a_{55} \end{pmatrix} \qquad (3.2)$$
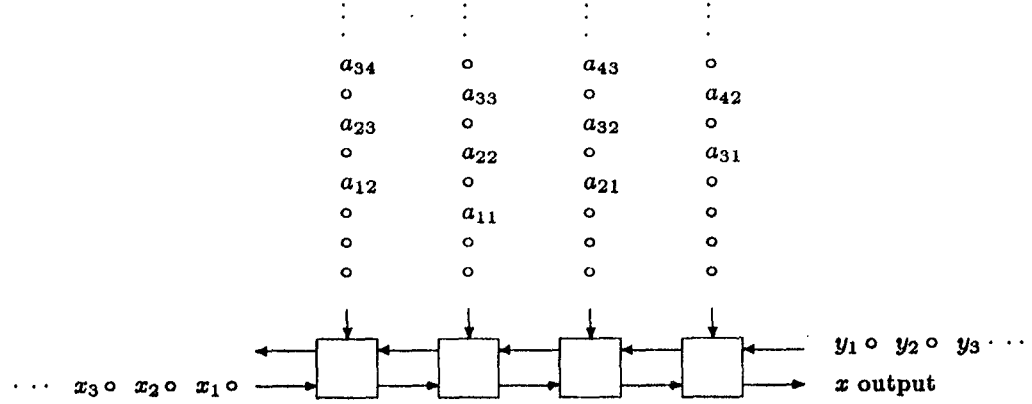

Figure 3.2.3. *Systolic layout for matrix vector multiplication.*

Figure 3.2.3 from (Kung & Leiserson [1980]) shows how to compute $y = Ax$ for A in (3.2). Each y_i is initially zero, and ⎯ stands for a one step delay. Figure 3.2.4 from (Kung & Leiserson [1980]), shows the array at several clock steps to show how the data flow and interconnections in Figure 3.2.3 calculate y.

For our example, A had bandwidth 3. In general, if A has bandwidth w then we need w IPPs to accomplish a matrix-vector multiplication. It takes $2n + w$ clock cycles for this algorithm to compute the matrix vector multiplication as compared to the sequential algorithm which takes $O(nw)$ units of time. As pointed out by Tables 3.2.1–3.2.2, a unit of time for the sequential algorithm is several clock cycles long. Therefore, the systolic algorithm is significantly faster.

The number of processors depends only on the bandwidth of the matrix A and not on its dimension. This makes a systolic design ideal for implementing *finite difference* approximations, where the number of mesh points is typically not known

a priori but where the maximum bandwidth is determined by the specific scheme. In the next section, we will show how the solution of (3.1) can be accomplished with systolic arrays.
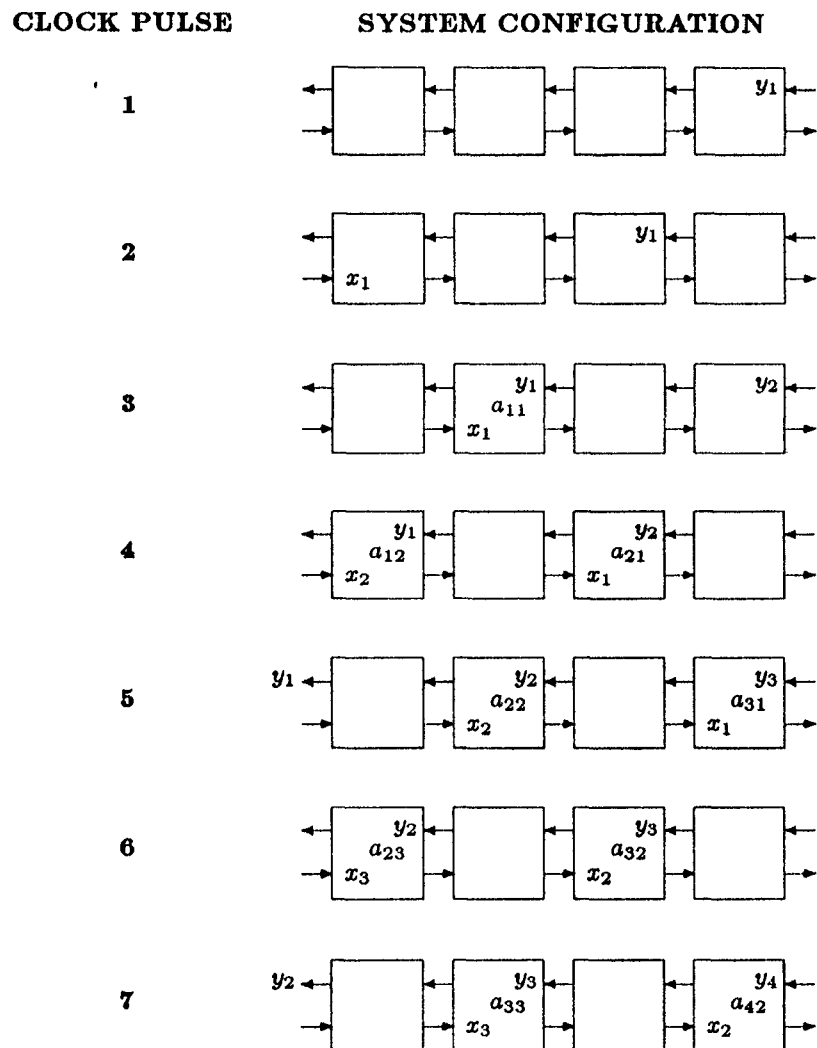
CLOCK PULSE   SYSTEM CONFIGURATION

1

2

3

4

5

6

7

Figure 3.2.4 *A walk through several clock cycles of the systolic layout.*

3.3 Results from Linear Algebra

To solve (3.1), we use Gaussian elimination without pivoting. This method, which is stable since $(I - \Delta t\, A_n)$ is strictly diagonally dominant, results in matrices L and U such that

$$LU = (I - \Delta t\, A_n)$$

where U is an upper triangular matrix and L is a unit lower triangular matrix.

The matrices L and U will be bi-diagonal since Gaussian elimination without pivoting is used. As is standard with Gaussian elimination, once the factors L and U are found, (3.1) is solved by finding x such that

$$Lx = D_k V^k \tag{3.3}$$

and then by solving

$$U V^{k+1} = x \tag{3.4}$$

to get V^{k+1}.

Let L be the lower triangular matrix resulting from the factorization of a band matrix. We are interested in solving $Lx = b$ for the vector x. That is, we want to solve

$$
\begin{pmatrix}
l_{11} & 0 & & & \\
l_{21} & l_{22} & 0 & & \\
l_{31} & l_{32} & l_{33} & 0 & \\
0 & l_{42} & l_{43} & l_{44} & \\
& \ddots & \ddots & \ddots & \ddots
\end{pmatrix}
\begin{pmatrix}
x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots
\end{pmatrix}
=
\begin{pmatrix}
b_1 \\ b_2 \\ b_3 \\ b_4 \\ \vdots
\end{pmatrix}
\tag{3.5}
$$

This is solved by a simple back substitution algorithm. It is well known that this can be accomplished by systolic arrays (Kung & Leiserson [1980]).

To see this, notice that the elements x_i can be computed by the following recursion:

$$y_i^{(1)} = 0,$$

$$y_i^{(k+1)} = y_i^{(k)} + l_{ik} x_k, \tag{3.6}$$

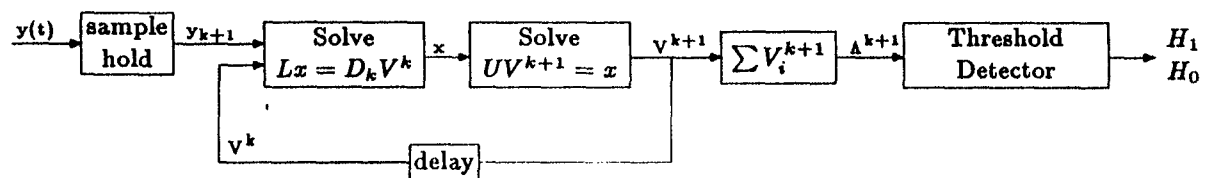$$x_i = (b_i - y_i^{(i)})/l_{ii}$$

shows the overall layout.

Figure 3.3.2. *Overall layout for the sequential detector.*
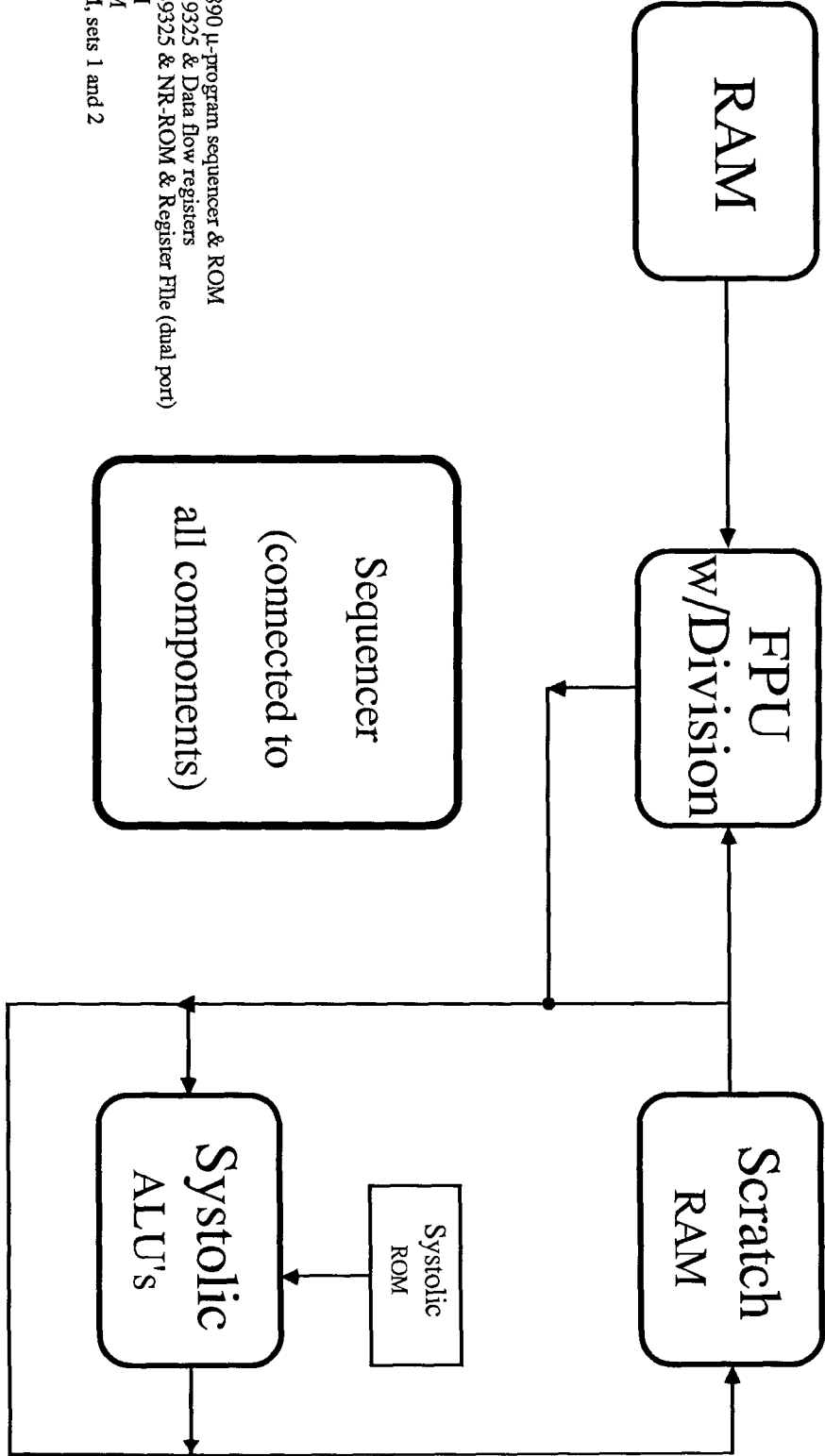
3.4 Implementation Issues

In collaboration with D. Simmons, a member of the research staff of the Systems Research Center, we have designed a board to implement the systolic architecture discussed above. This implementation will be completed shortly and will reside as a special purpose processor in an IBM personal computer. The hardware details of the *actual design* and *layout* will appear in a technical report. In order to get some idea of the implementation constraints, we will discuss some of the issues raised in going from the block diagram layout of Figure 3.3.2 to the hardware design.

The major implementation question was ' Should we use fixed point or floating point arithmetic?'. Among the advantages of fixed point arithmetic are that fixed point computations are up to four times faster than the corresponding floating point and that fixed point arithmetic leads to a simpler design. However, the numbers in our problems typically have large dynamic ranges. It is common to have density values as small as 1×10^{-6} with likelihood values as high as 1×10^{9}. In addition, we do not want to lock ourselves into a fixed dynamic range as would happen if we implemented fixed point. Therefore, in order to maintain flexibility, we decided to sacrifice processor speed and complexity and implement the design using IEEE standard floating point. A brief descriptioon of the processor layout, the architecture and the components of proposed board prototype design are shown in the next few pages. We have named the proposed chip the "Zakai I" chip. Once the processor is built we will be in a better position to evaluate the optimal design.
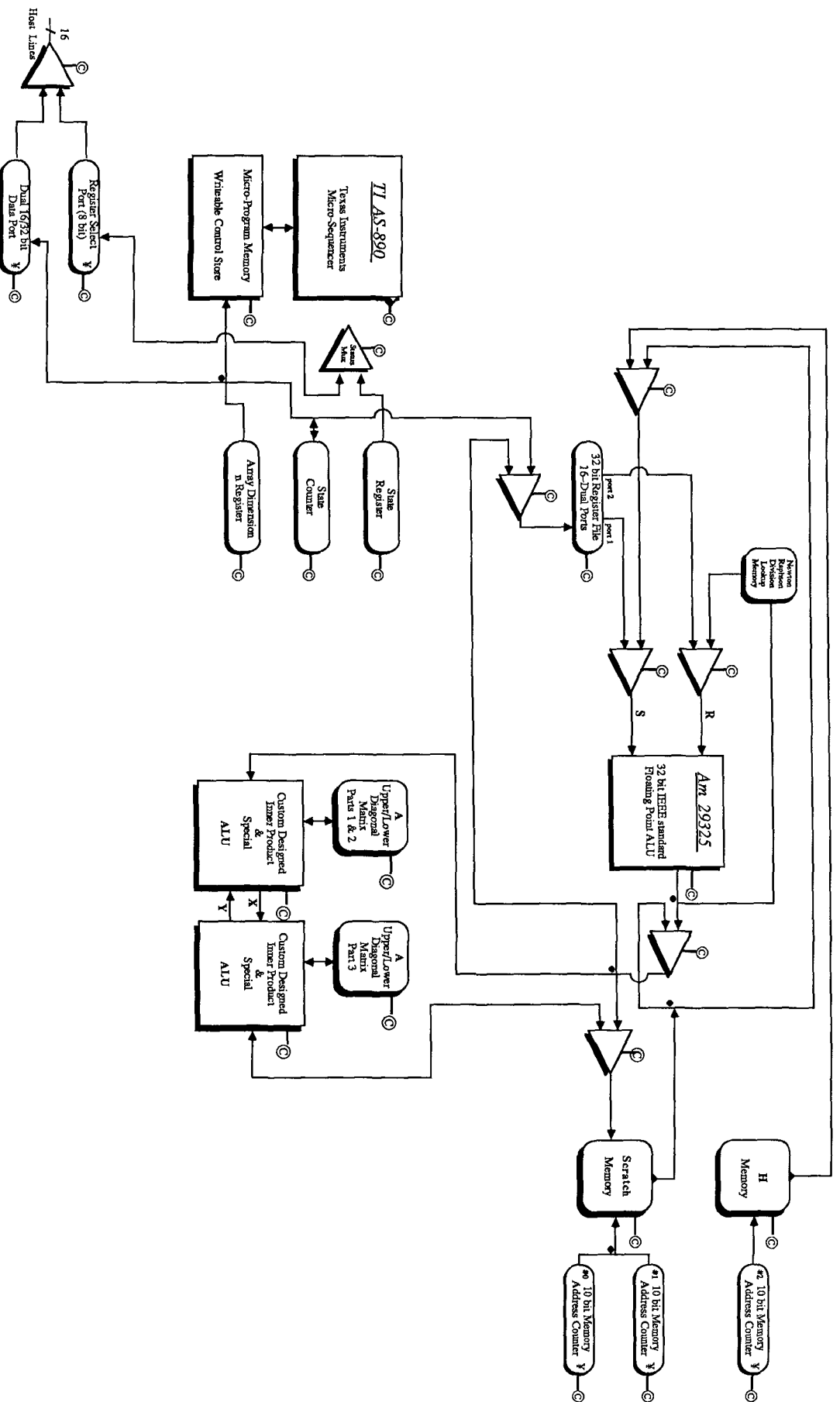
Initial timing calculations indicate a processor speed of 22 million floating point operations per second. This means that with 100 grid points, we can calculate 5000 solutions per second.
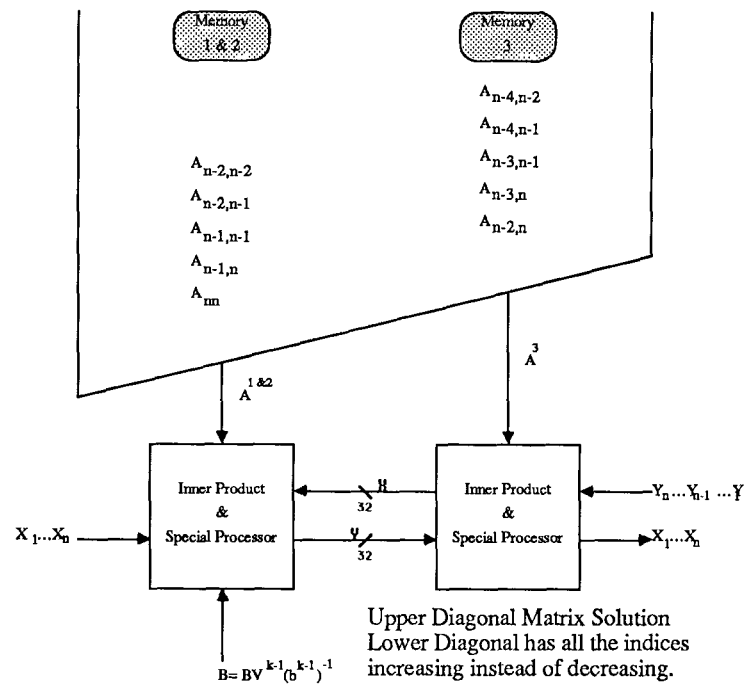
Zakai Solver Overview

RAM

FPU
w/Division

Sequencer
(connected to
all components)

Scratch
RAM

Systolic
ROM

Systolic
ALU's
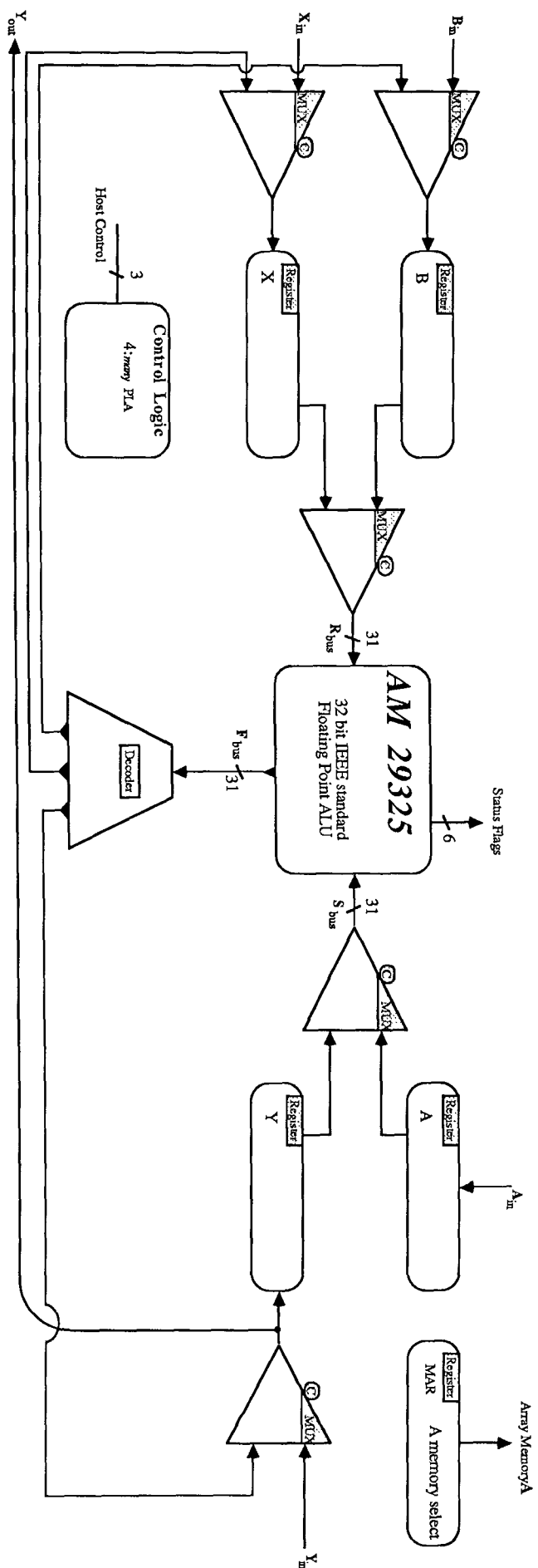
1. TI AS 890 μ-program sequencer & ROM
2. AMD 39325 & Data flow registers
3. AMD 39325 & NR-ROM & Register File (dual port)
4. S-RAM
5. H-RAM
6. A ROM, sets 1 and 2

Zakai 1 Solver Architecture

Actual Systolic Layout

Memory
1 & 2

Memory
3

$A_{n-4,n-2}$
$A_{n-4,n-1}$
$A_{n-2,n-2}$
$A_{n-3,n-1}$
$A_{n-2,n-1}$
$A_{n-3,n}$
$A_{n-1,n-1}$
$A_{n-2,n}$
$A_{n-1,n}$
A_{nn}

A^{3}

$A^{1 \& 2}$

| Inner Product & Special Processor | | Inner Product & Special Processor |

$X_1...X_n$

H
32

Y
32

$-Y_n...Y_{n-1}...Y$

$\blacktriangleright X_1...X_n$

$B = BV^{k-1}(b^{k-1})^{-1}$

Upper Diagonal Matrix Solution
Lower Diagonal has all the indices
increasing instead of decreasing.

Inner Product & Special Processor

3.5 Summary

We have shown that a systolic processor can be used to implement a sequential hypopaper tester. It was seen that the control structure for the resulting processor was simple and regular. This makes it ideal for VLSI implementation. Furthermore, timing results were given, hence exact computation time for the processor can be calculated.

This solves the sequential hypothesis testing problem for the scalar x case unfortunately more work remains for the vector x case.
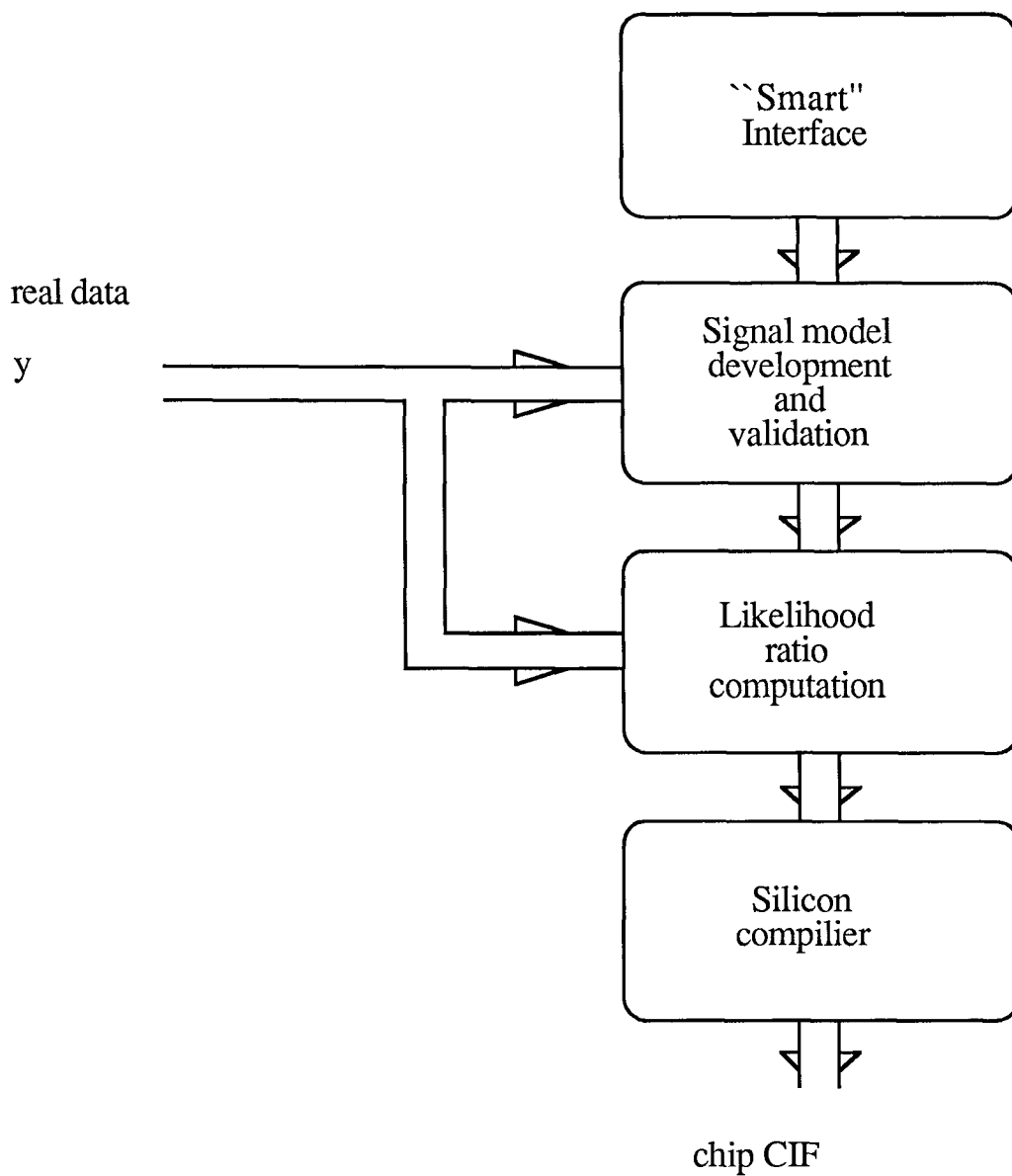
4. The DELPHI System

We are currently developing an expert system to facilitate the CAD of sophisticated and complex chips for non-linear signal processing. The software system is named DEsign Laboratory for Processing Hidden Information (DELPHI) and can be implemented on any AI machine carrying common LISP.

The system being developed combines an AI engine, symbolic algebra and multiprocessor numerical schemes. It has the capability of reasoning mathematically and makes available to a control engineer, sophisticated mathematical and computational tools, in a form suitable for immediate use. Its major advantage is its capability to interact symbolically with the user. The block diagram of the DELPHI software system is shown in figure 4.1 below. The system can be used as: (a) a tool for integrated design, (b) an advanced teaching aid (we have successfully used it with third year undergraduate students), (c) a tool for integrating symbolic and numerical computation.

The "smart" interface block, allows the user to enter a signal and observation model symbolically. We are currently implementing a module that will diagnose user "expertness", so as to allow various degrees of flexibility to the user. We are also implementing a module capable of understanding the "nature" of the module, i.e., diffusion, point or Markov chain type models.

The modelling block can currently build automatically a linear dynamical system model for a Guassian process using the Akaike-Rissanen-Hannan theory. We are

DELPHI Structure

Figure 4.1 Blocks of the DELPHI System

currently working on building general diffusion type, point process type and hidden Markov chain type models. This block will have the ability to call a variety of statistics and time series libraries from the AI machine for performing statistical validation tests on the model under construction.

The likelihood ratio (or sufficient statistic computer) block is under continuous development, and is currently the most advanced. In addition to the results of the present paper, it can perform similar computations for point process observations, mixed diffusion point process type models, and Markov chain models.

During next year we will couple the system to a silicon compiler for actual VLSI chip layout. What we will really develop is a "smart" compiler which can preselect the architecture to best fit the problem based on the *parallel* complexity theory of the Zakai equation.

The likelihood ratio block has currently the following capabilities: (an M indicates a MACSYMA computation, an F indicates a FORTRAN computation).

M1: Input f,g,h in symbolic form. Generate the multidimensional Zakai equation.

M2: Compute automatically discretizations of all stochastic differential equations. Automatically generate FORTRAN code.

F3: Solve numerically pathwise the resulting stochastic difference equations and store the $y.$ paths.

M4: Generate discretization schemes for the Zakai equation and automatically generate the associated FORTRAN code.

F5: Automatically integrate numerically the discretized Zakai equation.

F6: Display graphically the solution.

M7: Compute symbolically discretizations of the Likelihood Ratio and automatically generate appropriate FORTRAN code.

F8: Evaluate numerically the likelihood ratio.

Further additions currently planned include: apriori bounds for performance of

detectors and estimators, additional numerical schemes for the Zakai equation (e.g., multi-grid algorithms), automatic performance evaluation of sequential detectors. In addition we have initiated the development of a LISP based, domain-specific, higher level language for signal processing. Finally, a reduced version of the system is being ported on an IBM PC AT.

Future Directions

We have presented a VLSI design to implement a real time processor for the case where the signal process is scalar valued. More research work needs to be done in order to discover the appropriate architecture for the general case where the signal process is multidimensional. It is clear that this will be accomplished by exploiting problem specific information into the design. It is also clear that an implementable architecure will be very regular but exceedingly redundent and detailed. Therefore, a systematic and automatic method must be devised which takes a problem and produces a VLSI design.

It appears that systolic arrays can be used for higher dimensions, up to about 5 or 6. For even higher dimensions it is clear that we will need massively parallel architectures.

References

L. Arnold [1974], *Stochastic Differential Equations: Theory and Applications*, John Wiley & Sons, New York, NY.

J. S. Baras, G. L. Blankenship & W. E. Hopkins, Jr [1983], "Existence, uniqueness, and asymptotic behavior of solution to a class of Zakai equations with unbounded coefficients," *IEEE Trans. Auto. Control* AC-28, 203–214.

J.S. Baras [1981], "Approximate Solution of Nonlinear Filtering Problems by Direct Implementation of the Zakai Equation," *Proceedings of the 20th IEEE Conf. on Decision and Control*.

P. Besala [1975], "On the existence of a fundamental solution for a parabolic differential equation with unbounded coefficients," *Ann. Polonici Math.* 29, 403–409.

J. M. C. Clark [1978], "The design of robust approximations to the stochastic differential equations of nonlinear filtering," in *Communication Systems and Random Process Theory*, J. Shwirzynskii, ed., Sijthoff & Noordhoff , Alphen ann den Rijn–The Netherlands, 721–734.

M. H. A. Davis [1981], "PAthwise nonlinear filtering," in *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, M. Hasewinkel and J. C. Willems, ed., Reidel, Dordrect, Holland – Boston, Massachusetts – London, England, 505–528.

A. Friedman [1964], *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, NJ.

T. Kailath [1969], "A General Likelihood Ratio Formula for Random Signals in Gaussian Noise," *IEEE Trans. Inform. Theory* IT-15, 350–361.

T. Kato [1970], "Linear evolution equations of "hyperbolic" type," *J. Fac. Sci. Univ. Tokyo* 17, 241–258.

H. T. Kung & C. E. Leiserson [1980], "Algorithms for VLSI processor arrays," in *Introduction to VLSI Systems*, C. Mead and L. Conway, ed., Addison-Wesely, Reading, Massachusetts, 271–292.

H. J. Kushner [1977], *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*, Academic Press, New York, NY.

F. LeGland [1981], "Estimation de paramètres dans les processus stochastiques, en observation incomplète. Application à un problème de radio-astronomie," Thèse de Doct. Ing., Univ. Paris IX.

R. S. Liptser & A. N. Shiryayev [1977], *Statistics of Random Processes I : General Theory* (English Translation), Springer-Verlag, New York–Heidelberg–Berlin.

R. S. Liptser & A. N. Shiryayev [1978], *Statistics of Random Processes II : Applications* (English Translation), Springer-Verlag, New York–Heidelberg–Berlin.

E. Pardoux & D. Talay [1983], "Discretization and simulation of stochastic differential equations," in *Publication de Mathematiques Appliquees Marseille-Toulon*, Université de Provence, Marseille.

R. D. Richtmyer & K. W. Morton [1967], *Difference Methods for Initial-valued Problems* (Second Edition), Wiley-Interscience, New York, NY.

J. Schröder [1978], "M-matrices and generalizations using an operator theory approach," *SIAM Review* 20, 213–244.

A. N. Shiryayev [1977], *Optimal Stopping Rules* (English Translation), Springer-Verlag, New York–Heidelberg–Berlin.

H. F. Trotter [1958], "Approximation of semigroups of operators," *Pacific J. Math.* 8, 887–919.

Y. Yavin [1985], "Numerical Studies in Nonlinear Filtering," in *Lecture Notes in Control and Information Sciences 65*, A. V. Balakrishnan and M. Thoma, ed., Springer-Verlag, New York–Heidelberg–Berlin.

K. Yosida [1980], *Functional Analysis* (Sixth Edition), Springer-Verlag, New York–Heidelberg–Berlin.